

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2002, 01:04:31 ; Search time 62.27 Seconds

(without alignments)
242.589 Million cell updates/sec

Title: US-09-728-670-10

Perfect score: 715
Sequence: 1 SSSPDKGKXKGDASYEP.....SEHKNPGFNLTQVIEKK 136

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	136	17	AAW03071 Wild-type staphylo
2	715	100.0	136	17	AAW03086 Wild-type staphylo
3	715	100.0	136	22	AAW03086 Wild-type staphylo
4	715	100.0	136	22	AAW03086 Wild-type staphylo
5	715	100.0	136	22	AAW03086 Wild-type staphylo
6	712	99.6	136	20	AAW15024 Staphylokinase var
7	711	99.4	136	19	AAW44696 Staphylokinase var
8	711	99.4	136	20	AAW15022 Staphylokinase var
9	711	99.4	136	20	AAW15025 Staphylokinase var
10	711	99.4	136	20	AAW15016 Staphylokinase var
11	711	99.4	136	20	AAW15021 Staphylokinase var

12	711	99.4	136	22	AAW03071 standard; protein; 136 AA.
13	711	99.4	136	22	AAW03071 standard; protein; 136 AA.
14	711	99.4	136	22	AAW03071 standard; protein; 136 AA.
15	711	99.4	136	22	AAW03071 standard; protein; 136 AA.
16	710	99.3	136	20	AAW15013 Staphylokinase var
17	709	99.2	136	17	AAW03101 Staphylokinase der
18	709	99.2	136	20	AAW15023 Staphylokinase var
19	709	99.2	136	20	AAW15015 Staphylokinase var
20	709	99.2	136	20	AAW15010 Staphylokinase var
21	709	99.2	136	20	AAW15004 Staphylokinase var
22	709	99.2	136	20	AAW15006 Staphylokinase var
23	708	99.0	136	19	AAW44690 Staphylokinase mut
24	708	99.0	136	19	AAW44693 Staphylokinase mut
25	708	99.0	136	20	AAW15026 Staphylokinase var
26	708	99.0	136	20	AAW15027 Staphylokinase var
27	707	98.9	136	19	AAW44694 Staphylokinase mut
28	707	98.9	136	20	AAW15007 Staphylokinase var
29	707	98.9	136	20	AAW15018 Staphylokinase var
30	707	98.9	136	20	AAW15019 Staphylokinase var
31	707	98.9	136	20	AAW15011 Staphylokinase var
32	706	98.7	136	19	AAW44689 Staphylokinase mut
33	706	98.7	136	19	AAW44692 Staphylokinase mut
34	706	98.7	136	20	AAW15028 Staphylokinase var
35	706	98.7	136	20	AAW15014 Staphylokinase var
36	706	98.7	136	21	AAW01297 Wild type staphylo
37	705	98.6	136	19	AAW44691 Staphylokinase mut
38	705	98.6	136	19	AAW44695 Staphylokinase mut
39	704	98.5	136	20	AAW15041 Staphylokinase var
40	704	98.5	136	20	AAW15020 Staphylokinase var
41	704	98.5	136	20	AAW15012 Staphylokinase var
42	703	98.3	136	17	AAW03079 Staphylokinase der
43	703	98.3	136	17	AAW03078 Staphylokinase der
44	703	98.3	136	17	AAW03082 Staphylokinase der
45	703	98.3	136	17	AAW03076 Staphylokinase der

ALIGNMENTS

RESULT	1	AAW03071	standard; protein; 136 AA.
XX	AAW03071	standard; protein; 136 AA.	
AC	AAW03071	standard; protein; 136 AA.	
XX	AAW03071	standard; protein; 136 AA.	
DT	19-FEB-1997	(first entry)	
XX	19-FEB-1997	(first entry)	
DE	Wild-type staphylokinase.		
XX	Staphylokinase; mutant; mutagen; variant; immunogenicity; decrease; derivative; SakStav; arterial thrombosis; thrombolytic agent.		
KW	Staphylococcus aureus.		
XX	Staphylococcus aureus.		
OS	Staphylococcus aureus.		
XX	Staphylococcus aureus.		
FT	Key	Location/Qualifiers	
FT	Region	5..6	
FT	Region	/label= M20_epitope	
FT	Region	8..10	
FT	Region	/label= M21_epitope	
FT	Region	11..14	
FT	Region	/label= M1_epitope	
FT	Region	19	
FT	Region	/label= M22_epitope	
FT	Region	33..35	
FT	Region	/label= M2_epitope	
FT	Region	35..38	
FT	Region	/label= M3_epitope	
FT	Region	46..50	
FT	Region	/label= M4_epitope	
FT	Region	57..59	
FT	Region	/label= M5_epitope	
FT	Region	61..65	
FT	Region	/label= M6_epitope	

FT	Region	65..69	/label= M7_epitope
FT	Region	74..77	/label= M8_epitope
FT	Region	80..82	/label= M9_epitope
FT	Region	86..88	/label= M10_epitope
FT	Region	93..94	/label= M11_epitope
FT	Region	96..98	/label= M12_epitope
FT	Region	99..100	/label= M13_epitope
FT	Region	99..102	/label= M14_epitope
FT	Region	108..109	/label= M15_epitope
FT	Region	115..119	/label= M16_epitope
FT	Region	119..121	/label= M17_epitope
FT	Region	130	/label= M18_epitope
FT	Region	134..136	/label= M19_epitope
FT	Region		
PN	EP721982-A1.		
XX	17-JUL-1996.		
PD	06-JAN-1995;	95EP-0200023.	
XX	06-JAN-1995;	95EP-0200023.	
PR	(COLL-) COLLEN D J.		
XX	(LEUV-) LEUVEN RES & DEV VZW.		
PA	Colleen D;		
P1	WPI; 1996-322832/33.		
DR	New staphylokinase mutants with reduced immunogenicity - useful for		
XX	treating arterial thrombosis		
PT	Disclousure; Page 16; 21pp; English.		
XX			
PS	Staphylokinase derivatives showing a reduced immunogenicity as		
CC	compared to wild-type staphylokinase are claimed. The derivatives		
CC	are useful as thrombolytic agents to treat arterial thrombosis and		
CC	are pref. produced by eliminating at least one of the epitopes		
CC	indicated in the features table. The epitopes are destroyed by		
CC	replacing one or more amino acid residues in a charge cluster by an		
CC	Ala residue. Mutations are introduced using site-directed		
CC	mutagenesis on wild-type staphylokinase genomic DNA from the S. aureus		
CC	lysogenic strain SakSTAR. The present sequence is that of wild-type		
CC	SakSTAR staphylokinase.		
XX			
XX	Sequence 136 AA;		
XX			

	Query Match	100.0%	Score 715;	DB 17	Length 136;
	Best Local Similarity	100.0%	Pred. No. 1.9e-70;		
	Matches 136;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1	SSSFDKGKRYKKGDASFEPTGPLYLMVNTGVDSKGNELLSPHYVEPIKRGTTLKEKI	60		
Db	1	sssfcdkqkykkgdasfepitgplylmvntgydskgnellslaphyvefpikrgttlkeki	60		
OY	61	EYYEWALDNTAAKKEFRVVELDPSAKIEETTYVDKKRKEEKSPRIIEKGVPDDISEHI	120		
Db	61	eyyewaldntaakefrvvelcdpsakieevtyydkrkkkeekspriiekgvvpddisehi	120		

Oy	121	KNPGENITITKVTIEKK 136
Db	121	knpgfnlItkvIeKk 136
RESULT 2		
ID	AAW03086	
XX	AAW03086	standard; protein; 136 AA.
AC	AAW03086;	
XX	DT	19-FEB-1997 (first entry)
XX	DE	Wild-type staphylokinase.
XX		
KW		Staphylokinase; mutant; mutein; variant; immunogenicity; decrease
KM		derivative; SakSTAR; arterial thrombosis; thrombolytic agent.
XX	OS	Staphylococcus aureus.
XX		
FH	Key	Location/Qualifiers
FT	Region	5..6
FT		/label= M20_epitope
FT	Region	8..10
FT		/label= M21_epitope
FT	Region	11..14
FT		/label= M1_epitope
FT	Region	19
FT		/label= M22_epitope
FT	Region	33..35
FT		/label= M2_epitope
FT	Region	35..38
FT		/label= M3_epitope
FT	Region	46..50
FT		/label= M4_epitope
FT	Region	57..59
FT		/label= M5_epitope
FT	Region	61..65
FT		/label= M6_epitope
FT	Region	65..69
FT		/label= M7_epitope
FT	Region	74..77
FT		/label= M8_epitope
FT	Region	80..82
FT		/label= M9_epitope
FT	Region	86..88
FT		/label= M10_epitope
FT	Region	93..94
FT		/label= M11_epitope
FT	Region	96..98
FT		/label= M12_epitope
FT	Region	99..100
FT		/label= M13_epitope
FT	Region	99..102
FT		/label= M14_epitope
FT	Region	108..109
FT		/label= M15_epitope
FT	Region	115..119
FT		/label= M16_epitope
FT	Region	119..121
FT		/label= M17_epitope
FT	Region	130
FT		/label= M18_epitope
FT	Region	134..136
FT		/label= M19_epitope
XX		
PN	WO9621016-A2.	
XX		
PD	11-JUL-1996.	
XX		
PF	03-JAN-1996;	96WO-EP00081.
XX	17-NOV-1995;	95UP-0299781.

PR	06-JAN-1995;	95EP-0200023.
PR	11-JAN-1995;	95US-0371505.
PR	09-JUN-1995;	95EP-0201531.
PR	06-JUL-1995;	95US-0490992.
XX		
PA	(COLL.) COLLEN D.	
PA	(LEUV-) LEUVEN RES & DEV VZW.	
XX		
PI	COLLEN D:	
XX		
DR	WPI; 1996-333991/33.	
XX		
PT	New staphylokinase derivs. having reduced immunogenicity - useful	
PT	for treating arterial thrombosis	
XX		
PS	Disclosure; Fig 1; 58pp; English.	
XX		
CC	Staphylokinase derivatives showing a reduced immunogenicity as	
CC	compared to wild-type staphylokinase are claimed. The derivatives	
CC	are useful as thrombolytic agents to treat arterial thrombosis and	
CC	are pref. produced by eliminating at least one of the epitopes	
CC	indicated in the features table. The epitopes are destroyed by	
CC	replacing one or more amino acid residues in a charge cluster by an	
CC	Ala residue. Mutations are introduced using site-directed	
CC	mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus	
CC	lysogenic strain SakSTAR. The present sequence is that of wild-type	
XX	SakSTAR staphylokinase.	
XX		
Sequence	136 AA;	
50		

Query Match	100.0%	Score 715;	DB 17	Length 136;
Best Local Similarity	100.0%	Pred. No. 1.9e-70;		
Matches 136; Conservative	0;	Mismatches	0;	Gaps 0;

Qy	1	SSSDKGGYKKKKGDASAEPTGPIAMNMGVSKGHELLSPHVEEPIRPGTITAEKI	60
Qy	1	SSSDKGGYKKKKGDASAEPTGPIAMNMGVSKGHELLSPHVEEPIRPGTITAEKI	60
Db	1	ssfdkgykkgdaasyfepcgpylwmnvgvdsakgnellsphyvefpikpctltteki	60
Qy	61	EYYEMALDAAYKEPFAVELDPSAKTEMYNYDKNKKKEETKSPPTKEKFPVDDISEHI	120
Db	61	eyyewaldataayefrveladpsaklevcydnkknkkeetsrptlfekvfpvddlsehi	120
Qy	121	KNPGNLTITKVVIEKK	136
Qy	121	KNPGNLTITKVVIEKK	136
Db	121	knpgfnlttkvviekk	136

RESULT	3
AAB99421	
ID	AAB99421 standard; Protein; 136 AA

DT	28-AUG-2001 (first entry)
XX	
DE	Staphylokinase (SakSTAR) primary protein sequence.
XX	
KW	Staphylokinase; Staphylococcus aureus; SakSTAR; T-cell epitope;
KM	immunogenic; thrombolytic; acute myocardial infarction; immunogenicity
XX	
OS	Staphylococcus aureus.
XX	
PN	MO200140281-A2.
XX	
PD	07-JUN-2001.
XX	
PF	01-DEC-2000; 2000MO-EP12299.
XX	
PR	02-DEC-1999; 99EP-0204093.
XX	
PA	(THRO-) THROMB-X NV.
XX	

PI Warmerdam PAM, Plaisance SDNGH, Collen DJ, De Maeyer MCH,
XX
DR WPI; 2001-374786/39.

PT Reducing immunogenicity of protein (P) by eliminating T cell epitopes
PT in test peptides having amino acid sequence corresponding to (P) and
PT modifying amino acid sequence of (P) according to test peptide
PT modifications -

PS Disclosure; Fig 1; 50pp; English

CC The present invention describes a method for reducing the immunogenicity
CC of a peptide or protein (I), where (I) can be streptolysin (SakSTAR)
CC isolated from *Staphylococcus aureus*. The method involves designing a
CC series of overlapping test peptides having an amino acid sequence
CC corresponding to (I), modifying test peptides which are identified to
CC comprise one or more T-cell epitopes, such that they are reduced or
CC eliminated, and repeating the T-cell eliminating modifications for (I)
CC to produce a modified peptide or protein. Methods from the present
CC invention can be used for the treatment, diagnosis or prophylaxis of
CC for the preparation of a pharmaceutical composition for the treatment,
CC diagnosis or prophylaxis of a human subject. Streptolysin is used as a
CC potent thrombolytic agent in patients with acute myocardial infarction.
CC The method is useful for reducing cell based immunogenicity of non-human
CC proteins such as streptolysin or antibodies or their fragments, from
CC other species, for diagnostics and treatment of human disease. AAB99440
CC to AAB99449 represent amino acid sequences used in the exemplification
CC of the present invention.

sq Sequence 136 AA;

Query Match	100.0%	Score 715;	DB 22;	Length 136;
Best Local Similarity	100.0%	Pred. No. 1.9e-70;		
Matches 136; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

Oy	1	SSSEFGKGGKGGDASYPEPGLPMNVNVSCKGDELLSPHVEPPIRGTLTEKI	60	
		1	SSSfdgkykkdgaasyfepgtrylmnrvvgvskngnellspbyvetfipgltlcteki	60
Oy	61	EYYVMALDAAYAEFFRVLEDPDSAKITEMYYDNKKKKKEEKSPRTTEKGFVVPDLSEHI	120	
		61	eyyyewaldaaayaeffrvlepdpsaklevlyddknkkkeekslptekgfvvpdlseh1	120
Oy	121	KNPGFNLITTKVVIIEKK	136	
		121	knpgfnlittkvviiekk	136

RESULT	4
AAB61909	
ID	AAB61909 standard; Protein; 136 AA

DT	08-MAY-2001	(first entry)
XX		
DE	S. aureus staphylokinase G34S variant.	
XX		
KM	Immunogenicity: staphylokinase; variant; stability.	
XX		
OS	Staphylococcus aureus.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference	34
FT		/label= G34S
FT		/note= "wild-type Gly is replaced by Ser"
XX		
PN	W0200104287-A1.	
XX		
PD	18-JAN-2001.	
XX		
PF	06-JUL-2000; 2000WO-DK00371.	

```

XX 07-JUL-1993; 99DK-0000988.
PR 27-AUG-1999; 99DK-0001196.
PR 02-MAR-2000; 2000DK-0000339.
PR 18-MAY-2000; 2000DK-0000804.
XX (MAX-) MAYGEN ADS.
XX
XX Halkier T, Pedersen AH, Okkels JS;
XX
XX MPI: 2001-138342/14.
XX
XX Producing polypeptides with altered immunogenicity or improved
XX stability, comprising expressing a diversified nucleotide sequence
XX population and selecting polypeptides with altered immunogenicity or
XX improved stability -
XX
XX Example 1; Page 76; 83pp; English.
XX
XX The invention relates to a method of altering immunogenicity and/or
XX increasing stability of a polypeptide of interest. The method comprises
XX (a) expressing a diversified population of nucleotide sequences encoding
XX a polypeptide of interest; (b) screening the polypeptides expressed for
XX function, immunogenicity and/or stability; and (c) selecting functional
XX polypeptides with altered immunogenicity and/or increased stability. The
XX method is used to improve the properties of polypeptides, in particular
XX to alter the immunogenicity and/or increase the functional in vivo half-
XX life of the polypeptide. The method uses a high throughput system that
XX makes it possible to search several orders of magnitude more polypeptides
XX than is possible by previously known approaches. This enhances the chance
XX of finding the optimal variant from the many thousands of variants that
XX may be produced. The present sequence represents S. aureus staphylokinase
XX variant sequence, used to exemplify the method of the invention.
XX
XX Sequence 136 AA:
SQ

```

```

Query Match 100.0%; Score 715; DB 22; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.9e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 SSSFDKGYKKGGDASFEPTGPIYLMVNWGVDSKGNELSPHYVEFPPIKPGTTLTKRKI 60
DB 1 sssfdgkykkgddasyfepcpgylmwnvgvdsksgnellsphvefpikpgtltlckxi 60
QY 61 EYVEWALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
DB 61 eyvewaldataykefrvveldpsakievtydknkkkeetsfptekgfvpdlsehi 120
QY 121 KNPGFNLTITKVVIEKK 136
DB 121 knpgfnltitkviekk 136

```

```

RESULT 5
AAR39150
ID AAR39150 standard; Protein; 137 AA.
XX
XX AAR39150;
XX
XX 03-DEC-1993 (first entry)
XX
XX Staphylokinase SAK-STAR.
XX
XX Signal sequence, plasminogen activator; thrombosin; staphylokinase;
XX SAK.
XX
XX Staphylococcus aureus strain 23.
XX
XX MO9313209-A.
XX
XX 08-JUL-1993.
XX

```

```

PF 28-DEC-1992; 92WO-EP02989.
XX
XX 30-DEC-1991; 91DE-4143279.
PR 22-JUN-1992; 92DE-4220516.
PR 01-DEC-1992; 92DE-4240801.
XX
XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
XX
XX Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
XX
XX MPI: 1993-227325/28.
XX
XX N-PSDB; AAQ44270.
XX
XX Staphylo-kinase (SAK) sequences lacking sequences for signal
XX peptide(s) - for prodn. of proteins used as plasminogen
XX activators in thrombosis treatment, and monoclonal antibodies
XX against SAK
XX
XX Claim 3; Fig 3; 99pp; German.
XX
XX DNA encoding SAK lacking the signal peptide, is expressed
XX intracellularly. This avoids the problem of fast degradation of the
XX polypeptides or destruction of the host when expressed into the
XX medium or into the periplasm respectively. High expression is
XX possible and the chemically induced overprod. is easy to handle.
XX Also, the prods. are homogeneous.
XX SAK-polypeptide derivs. are plasminogen activators for the
XX treatment of thrombosis.
XX
XX Sequence 137 AA:
SQ

```

```

Query Match 100.0%; Score 715; DB 14; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.9e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 SSSFDKGYKKGGDASFEPTGPIYLMVNWGVDSKGNELSPHYVEFPPIKPGTTLTKRKI 60
DB 2 sssfdgkykkgddasyfepcpgylmwnvgvdsksgnellsphvefpikpgtltlckxi 61
QY 61 EYVEWALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
DB 62 eyvewaldataykefrvveldpsakievtydknkkkeetsfptekgfvpdlsehi 121
QY 121 KNPGFNLTITKVVIEKK 136
DB 122 knpgfnltitkviekk 137

```

```

RESULT 6
AAV15024
ID AAV15024 standard; Protein; 136 AA.
XX
XX AAV15024;
XX
XX 03-NOV-1999 (first entry)
XX
XX Staphylokinase variant SakSTAR (V132L).
XX
XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
XX SakSTAR (V132L); anti-thrombotic; fibrinolytic; cardiac; veterinary;
XX Staphylococcus aureus; myocardial infarction; arterial thrombosis;
XX monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
XX
XX Staphylococcus aureus.
XX
XX Synthetic.
XX
XX key Location/Qualifiers
XX
XX Misc-difference 132 /note= "Wild type Val is substituted by Leu"
XX
XX MO9940198-A2.
XX

```


PD 12-AUG-1999.
 XX
 PF 04-FEB-1999; 99MO-EP00748.
 XX
 PR 06-FEB-1998; 98EP-0200365.
 PR 04-FEB-1998; 98EP-0200323.
 XX
 PA (COLL/) COLLEN D J.
 PA (LEUV-) LEUVEN RES & DEV VZW.
 XX
 PI Collen DJ;
 XX
 DR WPI; 1999-508504/42.
 XX
 PT Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis
 XX
 PS Claim 7; Page -; 101pp; English.
 XX
 CC The present sequence is a specifically claimed Staphylokinase SAKSTAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SAKSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with peg to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 XX
 SQ Sequence 136 AA;

Query Match 99.6%; Score 712; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 4e-70;
 Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSFDKGYKKGDASYFEPTGPLYLWNVTVGVDSKGNELSPHYVEFPKPGTTLTKRKI 60
 DB 1 sssfdkgykkgdasyfeptgplylwnvvtgvdskgnellsphyvefpkpgtltlkeki 60
 QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTVYDKNKKKEETKSPITTEKGVVVDLSEHI 120
 DB 61 eyvewaldataykefrvveldp sakievtyydknkkkeetksfptekgvvpdlsehi 120
 QY 121 KNPGFNLTITKVVTEKR 136
 DB 121 knpgfnlittkvvlekk 136

RESULT 7
 AAM44696
 ID AAM44696 standard; protein: 136 AA.
 XX
 AC AAM44696;
 XX
 DT 01-MAY-1998 (first entry)
 XX
 DE Staphylokinase from Staphylococcus aureus.
 XX
 KW Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.
 OS Staphylococcus aureus.
 XX
 PN JP10028587-A.
 XX
 PD 03-FEB-1998.

PF 19-JUL-1996; 96JP-0208991.
 XX
 PR 19-JUL-1996; 96JP-0208991.
 XX
 PA (HONS) YAKULT HONSHA KK.
 XX
 DR WPI; 1998-162525/15.
 XX
 PT Staphylokinase mutants - useful as thrombolytic agents
 XX
 PS Example 1; Page 1; 9pp; Japanese.
 XX
 CC The present sequence represents staphylokinase (SAK). The invention
 CC relates to seven specifically claimed staphylokinase mutants which are
 CC useful as thrombolytic agents. The mutants are: (2) a SAK mutant (D41N)
 CC in which the 14th Asp in the SAK which comprises 136 amino acids is
 CC replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th Glu in SAK is
 CC replaced by Gln; (4) a SAK mutant (K113A) in which the 86th Lys in SAK
 CC is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th Lys in SAK
 CC is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th Glu in SAK
 CC is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th Asp in SAK
 CC is replaced by Glu, and (8) a SAK mutant (K136A) in which the 109th Lys
 CC in SAK is replaced by Ala.
 XX
 SQ Sequence 136 AA;

Query Match 99.4%; Score 711; DB 19; Length 136;
 Best Local Similarity 99.3%; Pred. No. 5.2e-70;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSSFDKGYKKGDASYFEPTGPLYLWNVTVGVDSKGNELSPHYVEFPKPGTTLTKRKI 60
 DB 1 sssfdkgykkgdasyfeptgplylwnvvtgvdskgnellsphyvefpkpgtltlkeki 60
 QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTVYDKNKKKEETKSPITTEKGVVVDLSEHI 120
 DB 61 eyvewaldataykefrvveldp sakievtyydknkkkeetksfptekgvvpdlsehi 120
 QY 121 KNPGFNLTITKVVTEKR 136
 DB 121 knpgfnlittkvvlekk 136

RESULT 8
 AAY15022
 ID AAY15022 standard; protein: 136 AA.
 XX
 AC AAY15022;
 XX
 DT 03-NOV-1999 (first entry)
 XX
 DE Staphylokinase variant SAKSTAR (K74Q).
 XX
 KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
 KW SAKSTAR (K74Q); anti-thrombotic; fibrinolytic; cardiant; veterinary;
 KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
 KW monoclonal antibody; polyethylene glycol; peg; plasma clearance.
 XX
 OS Staphylococcus aureus.
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT Misc-difference 74 /note- "Wild type Lys is substituted by Gln"
 XX
 PN WO940198-A2.
 XX
 PD 12-AUG-1999.
 XX
 PF 04-FEB-1999; 99MO-EP00748.
 XX
 PR 06-FEB-1998; 98EP-0200365.

PR 04-FEB-1998; 98EP-0200323.
 XX
 PA (COLL/) COLLEN D J.
 PA (LEUV-) LEUVEN RES & DEV VZW.
 XX
 PI Collen DJ;
 XX
 DR WPI: 1999-508504/42.
 XX
 PT Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis
 XX
 PS Claim 7; Page -: 101pp; English.
 XX
 CC The present sequence is a specifically claimed staphylokinase SakSTAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SakSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 XX
 SQ Sequence 136 AA;

Query Match 99.4%; Score 711; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 5.2e-70;
 Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSFDFGKRYKKGDASFEPTGPIYLMVNVGVDSKGNELSPHYVEFPKPGTTLTKETI 60
 DB 1 sssfdgkykkgddasyfepctgpylmvntgvdskgnellsphyvefpikpgtltltek1 60
 QY 61 EYVWEALDQATAYKEFRVVELDPSAKIEVYVDKNNKKEETKSPITREKGFVVPDLSEHI 120
 DB 61 eyvweladqataykefrveldpsakievelydknnkkeeetksfptekgfivvpdlsehi 120
 QY 121 KNPGFNLITRKVIEKK 136
 DB 121 knpgfnlitrkvielekk 136

RESULT 9
 AAY15025
 ID AAY15025 standard; Protein; 136 AA.
 XX
 AC AAY15025;
 XX
 DT 03-NOV-1999 (first entry)
 XX
 DE Staphylokinase variant SakSTAR (V132T).
 XX
 KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
 KW SakSTAR (V132T); anti-thrombotic; fibrinolytic; cardiant; veterinary;
 KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
 KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
 XX
 OS Staphylococcus aureus.
 OS Synthetic.
 XX
 FH key location/Qualifiers
 FT Misc-difference 132 /note- "wild type Val is substituted by Thr"
 XX
 PN MO9940198-A2.

XX
 PD 12-AUG-1999.
 XX
 PE 04-FEB-1999; 99WO-EP00748.
 XX
 PR 06-FEB-1998; 98EP-0200365.
 PR 04-FEB-1998; 98EP-0200323.
 XX
 PA (COLL/) COLLEN D J.
 PA (LEUV-) LEUVEN RES & DEV VZW.
 XX
 PI Collen DJ;
 XX
 DR WPI: 1999-508504/42.
 XX
 PT Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis
 XX
 PS Claim 7; Page -: 101pp; English.
 XX
 CC The present sequence is a specifically claimed Staphylokinase SakSTAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SakSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 XX
 SQ Sequence 136 AA;

Query Match 99.4%; Score 711; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 5.2e-70;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSFDFGKRYKKGDASFEPTGPIYLMVNVGVDSKGNELSPHYVEFPKPGTTLTKETI 60
 DB 1 sssfdgkykkgddasyfepctgpylmvntgvdskgnellsphyvefpikpgtltltek1 60
 QY 61 EYVWEALDQATAYKEFRVVELDPSAKIEVYVDKNNKKEETKSPITREKGFVVPDLSEHI 120
 DB 61 eyvweladqataykefrveldpsakievelydknnkkeeetksfptekgfivvpdlsehi 120
 QY 121 KNPGFNLITRKVIEKK 136
 DB 121 knpgfnlitrkvielekk 136

RESULT 10
 AAY15016
 ID AAY15016 standard; Protein; 136 AA.
 XX
 AC AAY15016;
 XX
 DT 03-NOV-1999 (first entry)
 XX
 DE Staphylokinase variant SakSTAR (V132A).
 XX
 KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
 KW SakSTAR (V132A); anti-thrombotic; fibrinolytic; cardiant; veterinary;
 KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
 KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
 XX
 OS Staphylococcus aureus.
 OS Synthetic.

XX Key Location/Qualifiers
 FH Misc-difference 132 /note= "Wild type Val is substituted by Ala"
 FT
 XX
 PN W09940198-A2.
 XX
 PD 12-AUG-1999.
 XX
 PE 04-FEB-1999; 99WO-EP00748.
 XX
 PR 06-FEB-1998; 98EP-0200365.
 PR 04-FEB-1998; 98EP-0200323.
 XX
 PA (COLL/) COLLEN D J.
 PA (LEUV-) LEUVEN RES & DEV VZW.
 XX
 PI Collen DJ;
 DR WPI; 1999-508504/42.
 XX
 PT Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis
 XX
 PS Claim 7; Page -: 101pp; English.
 XX
 CC The present sequence is a specifically claimed Staphylokinase SakSTAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SakSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 XX
 SQ Sequence 136 AA;
 Query Match 99.4%; Score 711; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 5.2e-70;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SSSFDKGYKKGDADSYFEPTGPIYLVNVTGVDKGNELSPHYVEPIKPGTTLTKETI 60
 DB 1 sssfdkgykkgdasyfeptgpylmvntvgvdskgnellsphyvefpikpgttltkexi 60
 QY 61 EYVWMAIDATAYKERRVAVLDPSAKIEVTYYDKNKKKEKSPITTEKGFVPPDLSERI 120
 DB 61 eyvwaldataaykefrvavldpsakievtyydknkkkeetksfptekgfvpdpdlselh 120
 QY 121 KNPGENLITKVVIEKK 136
 DB 121 knpgfnlitrkvviekk 136
 RESULT 11
 AAY15021
 ID AAY15021 standard; Protein: 136 AA.
 AC AAY15021;
 DT 03-NOV-1999 (first entry)
 DE Staphylokinase variant SakSTAR (K35E).
 KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;

KW SakSTAR (K35E); anti-thrombotic; fibrinolytic; cardiant; veterinary;
 KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
 KM monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
 OS Staphylococcus aureus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 35 /note= "Wild type Lys is substituted by Glu"
 XX
 PN W09940198-A2.
 XX
 PD 12-AUG-1999.
 XX
 PE 04-FEB-1999; 99WO-EP00748.
 XX
 PR 06-FEB-1998; 98EP-0200365.
 PR 04-FEB-1998; 98EP-0200323.
 XX
 PA (COLL/) COLLEN D J.
 PA (LEUV-) LEUVEN RES & DEV VZW.
 XX
 PI Collen DJ;
 DR WPI; 1999-508504/42.
 XX
 PT Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis
 XX
 PS Claim 7; Page -: 101pp; English.
 XX
 CC The present sequence is a specifically claimed staphylokinase SakSTAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SakSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 XX
 SQ Sequence 136 AA;
 Query Match 99.4%; Score 711; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 5.2e-70;
 Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSSFDKGYKKGDADSYFEPTGPIYLVNVTGVDKGNELSPHYVEPIKPGTTLTKETI 60
 DB 1 sssfdkgykkgdasyfeptgpylmvntvgvdskgnellsphyvefpikpgttltkexi 60
 QY 61 EYVWMAIDATAYKERRVAVLDPSAKIEVTYYDKNKKKEKSPITTEKGFVPPDLSERI 120
 DB 61 eyvwaldataaykefrvavldpsakievtyydknkkkeetksfptekgfvpdpdlselh 120
 QY 121 KNPGENLITKVVIEKK 136
 DB 121 knpgfnlitrkvviekk 136
 RESULT 12
 AAB61908
 ID AAB61908 standard; Protein: 136 AA.
 AC AAB61908;

```

XX 08-MAY-2001 (first entry)
DT S. aureus staphylokinase mature protein sequence.
DE Immunogenicity; staphylokinase; variant; stability.
XX Staphylococcus aureus.
XX WO200104287-A1.
XX 18-JAN-2001.
XX 06-JUL-2000; 2000MO-DK00371.
XX 07-JUL-1999; 99DK-0000988.
XX 27-AUG-1999; 99DK-0001196.
XX 02-MAR-2000; 2000DK-0000339.
XX 18-MAY-2000; 2000DK-0000804.
XX (MAXY-) MAXYGEN APS.
XX Halkier T, Pedersen AH, Okkels JS;
XX WPI: 2001-138342/14.
XX N-PSDB; AAC85112.
XX Producing polypeptides with altered immunogenicity or improved
XX stability, comprises expressing a diversified nucleotide sequence
XX population and selecting polypeptides with altered immunogenicity or
XX improved stability.
XX Example 1; Page 75; 83pp; English.
XX The invention relates to a method of altering immunogenicity and/or
XX increasing stability of a polypeptide of interest. The method comprises
XX (a) expressing a diversified population of nucleotide sequences encoding
XX a polypeptide of interest; (b) screening the polypeptides expressed for
XX function, immunogenicity and/or stability; and (c) selecting functional
XX polypeptides with altered immunogenicity and/or increased stability. The
XX method is used to improve the properties of polypeptides, in particular
XX to alter the immunogenicity and/or increase the functional in vivo half-
XX life of the polypeptide. The method uses a high throughput system that
XX makes it possible to search several orders of magnitude more polypeptides
XX than is possible by previously known approaches. This enhances the chance
XX of finding the optimal variant from the many thousands of variants that
XX may be produced. The present sequence represents S. aureus staphylokinase
XX mature protein sequence, used to exemplify the method of the invention.
XX Sequence 136 AA:
SO

```

```

Query Match 99.4%; Score 711; DB 22; Length 136;
Best Local Similarity 99.3%; Pred. No. 5.2e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 SSSFDGKTKKGGDASYFEPTGYLMVNTGVDKGNELSPHYVEPIKPGTTLTKETI 60
   |||||||
DB 1 sssfdgkykkgddasyfepitgpylmvntvgdgkgnellsphyvefipikgtltlkek1 60
OY 61 EYVEWALDATALYKEFRVVELDPSAKIEVYVDKNNKKKEETKSPITEKGFVVPDLSEHI 120
   |||||||
DB 61 eyvewaldatalykefrveldpsakievlydknkkkeetksfpitekgfvvpdlsehi 120
OY 121 KNPGFNLTITKVIEKK 136
   |||||||
DB 121 knpgfnltitkviekk 136

```

```

RESULT 13
AAR39149
ID AAR39149 standard; Protein: 137 AA.
XX

```

```

AC AAR39149;
XX 03-DEC-1993 (first entry)
DT Staphylokinase SAK-CphLC.
DE Signal sequence, plasminogen activator; thrombostis; staphylokinase;
XX SAK.
XX Staphylococcus aureus phage phiC.
XX WO9313209-A.
XX 08-JUL-1993.
XX 28-DEC-1992; 92MO-EP02989.
XX 30-DEC-1991; 91DE-4143279.
XX 22-JUN-1992; 92DE-4220516.
XX 01-DEC-1992; 92DE-4240801.
XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
XX Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
XX WPI: 1993-227325/20.
XX N-PSDB; AAQ44269.
XX Staphylo-kinase (SAK) sequences lacking sequences for signal
XX peptide(s) - for prodn. of proteins used as plasminogen
XX activators in thrombostis treatment, and monoclonal antibodies
XX against SAK
XX Claim 3; Fig 2; 99pp; German.
XX DNA encoding SAK lacking the signal peptide, is expressed
XX intracellularly. This avoids the problem of fast degradation of the
XX polypeptides or destruction of the host when expressed into the
XX medium or into the periplasm respectively. High expression is
XX possible and the chemically induced overprodn. is easy to handle.
XX Also, the prods. are homogeneous.
XX SAK-polypeptide derivs. are plasminogen activators for the
XX treatment of thrombostis.
XX Sequence 137 AA:
SO

```

```

Query Match 99.4%; Score 711; DB 14; Length 137;
Best Local Similarity 99.3%; Pred. No. 5.3e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 SSSFDGKTKKGGDASYFEPTGYLMVNTGVDKGNELSPHYVEPIKPGTTLTKETI 60
   |||||||
DB 2 sssfdgkykkgddasyfepitgpylmvntvgdgkgnellsphyvefipikgtltlkek1 61
OY 61 EYVEWALDATALYKEFRVVELDPSAKIEVYVDKNNKKKEETKSPITEKGFVVPDLSEHI 120
   |||||||
DB 62 eyvewaldatalykefrveldpsakievlydknkkkeetksfpitekgfvvpdlsehi 121
OY 121 KNPGFNLTITKVIEKK 136
   |||||||
DB 122 knpgfnltitkviekk 137

```

```

RESULT 14
AAR12137
ID AAR12137 standard; Protein: 163 AA.
XX AAR12137;
XX 05-AUG-1991 (first entry)
XX S.aureus Staphylokinase with signal peptide.
DE

```

```

XX SAK; protein production.
KW Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein /label= signal peptide
FT 28..163
FT /label= staphylokinase
XX
XX JP03098595-A.
XX
XX 24-APR-1991.
XX
XX 11-SEP-1989; 89JP-0234874.
XX
XX 11-SEP-1989; 89JP-0234874.
XX
XX (TAIS ) TAISHO PHARMACEUT KK.
XX
XX WPI; 1991-167039/23.
XX
XX N-PSDB; AAQ11813.
XX
XX Prep'n. of peptide(s) - by construction of expression
XX vector, transformation of E.coli etc., culturing to secrete
XX peptide(s) and collecting peptide(s)
XX
XX Example; Fig 3; 15pp; Japanese.
XX
XX The first 79 residues of this sequence form part of a fusion
XX protein with somatomedin C. A synthetic construct encoding the
XX SAK-SMC fusion is used to transform E.coli or Bacillus subtilis.
XX The SAK signal peptide directs extracellular secretion of the
XX somatomedin C.
XX See also AAQ11814.
XX
XX Sequence 163 AA:
SQ

```

Query Match 99.4%; Score 711; DB 12; Length 163;
 Best Local Similarity 99.3%; Pred. No. 6,7e-70;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 SSSFDKGRKKKGDDASYFEPTGYLMVNTGVDSKGNELSPHYVEFPKPGTTLTKRKI 60
   |||||||
DB 28 sssfdkgykkgddasyfeptgpylmvntgvdgkgnellsphyvefpkpgtlttkrk1 87
   |||||||
QY 61 EYVEMALDATALKEFRVVELDPSAKIEVTYIDKNNKKKEETKSFTTEKGFVPPDLSH 120
   |||||||
DB 88 eyvewaldataykefrvveldpsakievtyidknkkkeetksfptekegfvpdlsh1 147
   |||||||
QY 121 KNPGFNLITKRVIEKK 136
   |||||||
DB 148 knpgfnlltkvvlekk 163

```

RESULT 15
 AAR28844
 ID AAR28844 standard; Protein; 163 AA.
 AC AAR28844;
 XX
 XX 13-JAN-1993 (first entry)
 XX
 XX Staphylokinase (SAK).
 XX
 XX Staphylokinase; thrombosis; fibrinogen; plasminogen.
 XX
 XX Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..27

```

FT /label= sig_peptide
FT Protein 28..163
FT /label= mat_protein
XX
XX WO9211356-A.
XX
XX 09-JUL-1992.
XX
XX 17-DEC-1991; 91WO-JP01722.
XX
XX 17-DEC-1990; 90JP-0411063.
XX
XX (HONS ) YAKULT HONSHA KK.
XX
XX Hashimoto S, Matsumoto T, Matsuo O, Onoue M, Sakai M;
XX Sako T, Sansawa H, Shimura K, Shishido Y, Watanabe T;
XX Yokokura T;
XX
XX WPI; 1992-250079/30.
XX
XX Thrombolytic peptide SAK-11 derived from staphylokinase - useful
XX as a plasminogen activator for treating thrombosis
XX
XX Disclosure; Page 16; 26pp; Japanese.
XX
XX SAK-11 (AAR25468) is derived from its precursor, staphylokinase (SAK)
XX - represented in AAR28844, by tryptic cleavage of its 10 N-terminal
XX amino acid residues. SAK-11 is useful as the active agent in an
XX injectable treatment for thrombosis. Thrombolytic activity is
XX demonstrated using a rabbit jugular vein thrombosis model. A clear
XX decrease in fibrinogen is observed upon application of 0.9 mg SAK-11/
XX kg. Plasminogen activation is superior to that of SAK (no specific
XX data given). Furthermore, SAK-11 has very low antigenicity as shown
XX in a PCA study, against streptokinase and egg albumin controls, in
XX BALB/c mice.
XX
XX Sequence 163 AA:
SQ

```

Query Match 99.4%; Score 711; DB 13; Length 163;
 Best Local Similarity 99.3%; Pred. No. 6,7e-70;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 SSSFDKGRKKKGDDASYFEPTGYLMVNTGVDSKGNELSPHYVEFPKPGTTLTKRKI 60
   |||||||
DB 28 sssfdkgykkgddasyfeptgpylmvntgvdgkgnellsphyvefpkpgtlttkrk1 87
   |||||||
QY 61 EYVEMALDATALKEFRVVELDPSAKIEVTYIDKNNKKKEETKSFTTEKGFVPPDLSH 120
   |||||||
DB 88 eyvewaldataykefrvveldpsakievtyidknkkkeetksfptekegfvpdlsh1 147
   |||||||
QY 121 KNPGFNLITKRVIEKK 136
   |||||||
DB 148 knpgfnlltkvvlekk 163

```

RESULT 16
 AAY15013
 ID AAY15013 standard; Protein; 136 AA.
 AC AAY15013;
 XX
 XX 03-NOV-1999 (first entry)
 XX
 XX Staphylokinase variant SakstrAR (T90A).
 XX
 XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
 KW SakstrAR (T90A); anti-thrombotic; fibrinolytic; cardiac; veterinary;
 KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
 KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
 OS
 OS Staphylococcus aureus.
 OS Synthetic.


```

XX AAY15023:
AC 03-NOV-1999 (first entry)
DT XX
DE Staphylokinase variant SakSTAR (K130T).
XX
KM Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KM SakSTAR (K130T); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KM Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KM monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
XX
OS Staphylococcus aureus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 130 /note= "Wild type Lys is substituted by Thr"
XX
XX MO9940198-A2.
XX 12-AUG-1999.
XX
XX 04-FEB-1999; 99WO-EP00748.
XX
XX 06-FEB-1998; 98EP-0200365.
XX 04-FEB-1998; 98EP-0200323.
XX
XX (COLL/) COLLEN D. J.
XX (LEUV-) LEUVEN RES & DEV VZW.
XX
XX Collen DJ:
XX
XX WPI; 1999-508504/42.
XX
XX Staphylokinase derivatives with reduced immunogenicity, used for,
XX e.g. treatment of arterial thrombosis
XX
XX Claim 7; Page -: 101pp; English.
XX
XX The present sequence is a specifically claimed Staphylokinase SakSTAR
XX variant. This variant has one aminoacid that has been substituted
XX by another aminoacid that reduces the reactivity with monoclonal
XX antibodies and absorption of SakSTAR-specific antibodies from plasma
XX of patients treated with staphylokinase. The derivatives can also be
XX substituted with cysteine modified with PEG to maintain the specific
XX activity and significantly reduce the plasma clearance. They have
XX altered immunogenicity without markedly reducing the specific activity.
XX This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
XX potency in human plasma. The new staphylokinase derivatives are used
XX for treatment of arterial thrombosis, especially myocardial infarction.
XX The compositions can be used in human or veterinary practice.
XX Note: The present sequence is not shown in the specification, but is
XX derived from the Staphylococcus aureus wild type staphylokinase sequence
XX given in figure 1.
XX
XX Sequence 136 AA;
SQ

```

Query Match 99.2%; Score 709; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 8.6e-70;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 SSSFDKGYKKGGDDASYFEPTGPIYIMVNVGVDSKGNELLSPHYVEFPKRGTTLTKEKI 60
DB 1 SSSIDKGYKKGGDDASYFEPTGPIYIMVNVGVDSKGNELLSPHYVEFPKRGTTLTKEKI 60
OY 61 EYVEMALDATAKKEFRVVELDPSAKIEVYTYDKNKKKKEETKSFEITKRGVVDLSSEHI 120
DB 61 EYVEMALDATAKKEFRVVELDPSAKIEVYTYDKNKKKKEETKSFEITKRGVVDLSSEHI 120
OY 121 KNGGNLITTKVIEKK 136
DB 121 KNGGNLITTKVIEKK 136

```

```

DB 121 knggnlittvlekk 136
RESULT 19
AAY15015
ID AAY15015 standard; Protein; 136 AA.
XX
XX AAY15015;
AC
XX
XX 03-NOV-1999 (first entry)
DT
XX
DE Staphylokinase variant SakSTAR (K130A).
XX
XX
KM Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KM SakSTAR (K130A); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KM Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KM monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
XX
OS Staphylococcus aureus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 130 /note= "Wild type Lys is substituted by Ala"
XX
XX MO9940198-A2.
XX 12-AUG-1999.
XX
XX 04-FEB-1999; 99WO-EP00748.
XX
XX 06-FEB-1998; 98EP-0200365.
XX 04-FEB-1998; 98EP-0200323.
XX
XX (COLL/) COLLEN D. J.
XX (LEUV-) LEUVEN RES & DEV VZW.
XX
XX Collen DJ:
XX
XX WPI; 1999-508504/42.
XX
XX Staphylokinase derivatives with reduced immunogenicity, used for,
XX e.g. treatment of arterial thrombosis
XX
XX Claim 7; Page -: 101pp; English.
XX
XX The present sequence is a specifically claimed Staphylokinase SakSTAR
XX variant. This variant has one aminoacid that has been substituted
XX by another aminoacid that reduces the reactivity with monoclonal
XX antibodies and absorption of SakSTAR-specific antibodies from plasma
XX of patients treated with staphylokinase. The derivatives can also be
XX substituted with cysteine modified with PEG to maintain the specific
XX activity and significantly reduce the plasma clearance. They have
XX altered immunogenicity without markedly reducing the specific activity.
XX This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
XX potency in human plasma. The new staphylokinase derivatives are used
XX for treatment of arterial thrombosis, especially myocardial infarction.
XX The compositions can be used in human or veterinary practice.
XX Note: The present sequence is not shown in the specification, but is
XX derived from the Staphylococcus aureus wild type staphylokinase sequence
XX given in figure 1.
XX
XX Sequence 136 AA;
SQ

```

Query Match 99.2%; Score 709; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 8.6e-70;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 SSSFDKGYKKGGDDASYFEPTGPIYIMVNVGVDSKGNELLSPHYVEFPKRGTTLTKEKI 60
DB 1 SSSIDKGYKKGGDDASYFEPTGPIYIMVNVGVDSKGNELLSPHYVEFPKRGTTLTKEKI 60
OY 61 EYVEMALDATAKKEFRVVELDPSAKIEVYTYDKNKKKKEETKSFEITKRGVVDLSSEHI 120
DB 61 EYVEMALDATAKKEFRVVELDPSAKIEVYTYDKNKKKKEETKSFEITKRGVVDLSSEHI 120
OY 121 KNGGNLITTKVIEKK 136
DB 121 KNGGNLITTKVIEKK 136

```

```

QY 61 EYVEWALDATTAKKEFVVELDPSAKIEVITYYDKNKKKEETKSPRTTEKGFVVPDLSEHI 120
DB 61 EYVEWALDATTAKKEFVVELDPSAKIEVITYYDKNKKKEETKSPRTTEKGFVVPDLSEHI 120
QY 121 KNPGFNLTITKVIIEKK 136
DB 121 KNPGFNLTITKVIIEKK 136

RESULT 20
ID AAY15010 standard; protein; 136 AA.
AC AAY15010:
XX
XX
XX 03-NOV-1999 (first entry)
DE Staphylokinase variant SakSTAR (K35A).
KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KW SakSTAR (K35A); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
OS Staphylococcus aureus.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FT MISC-difference 35
XX FT /note= "Wild type Lys is substituted by Ala"
XX
XX PN W09940198-A2.
XX PD 12-AUG-1999.
XX PF 04-FEB-1999; 99MO-EP00748.
XX PR 06-FEB-1998; 98EP-0200365.
XX PR 04-FEB-1998; 98EP-0200323.
XX
XX PA (COLL/) COLLEN D J.
XX PA (LEUV-) LEUVEN RES & DEV VZW.
XX
XX PI Collen DJ.
XX
XX DR WPI; 1999-508504/42.
XX
XX PT Staphylokinase derivatives with reduced immunogenicity, used for,
XX PT e.g. treatment of arterial thrombosis
XX
XX PS Claim 7; Page -: 101pp; English.
XX
XX CC The present sequence is a specifically claimed Staphylokinase SakSTAR
XX CC variant. This variant has one aminoacid that has been substituted
XX CC by another aminoacid that reduces the reactivity with monoclonal
XX CC antibodies and absorption of SakSTAR-specific antibodies from plasma
XX CC of patients treated with staphylokinase. The derivatives can also be
XX CC substituted with cysteine labelled with PEG to maintain the specific
XX CC activity and significantly reduce the plasma clearance. They have
XX CC altered immunogenicity without markedly reducing the specific activity.
XX CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
XX CC potency in human plasma. The new staphylokinase derivatives are used
XX CC for treatment of arterial thrombosis, especially myocardial infarction.
XX CC The compositions can be used in human or veterinary practice.
XX CC Note: The present sequence is not shown in the specification, but is
XX CC derived from the Staphylococcus aureus wild type staphylokinase sequence
XX CC given in figure 1.
XX
XX SX Sequence 136 AA;

```

```

Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 SSSFDKGRYKGGDASYEPRGPLYMNVGVDSKGNELSPHYVEPPIKGTTLTKRKI 60
DB 1 SSSFDKGRYKGGDASYEPRGPLYMNVGVDSKGNELSPHYVEPPIKGTTLTKRKI 60
QY 61 EYVEWALDATTAKKEFVVELDPSAKIEVITYYDKNKKKEETKSPRTTEKGFVVPDLSEHI 120
DB 61 EYVEWALDATTAKKEFVVELDPSAKIEVITYYDKNKKKEETKSPRTTEKGFVVPDLSEHI 120
QY 121 KNPGFNLTITKVIIEKK 136
DB 121 KNPGFNLTITKVIIEKK 136

RESULT 21
ID AAY15004 standard; protein; 136 AA.
AC AAY15004:
XX
XX
XX 03-NOV-1999 (first entry)
DE Staphylokinase variant SakSTAR (E75A).
KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KW SakSTAR (E75A); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance;
XX veterinary.
XX
XX OS Staphylococcus aureus.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT MISC-difference 75
XX FT /note= "Wild type Glu is substituted by Ala"
XX
XX PN W09940198-A2.
XX PD 12-AUG-1999.
XX PF 04-FEB-1999; 99MO-EP00748.
XX PR 06-FEB-1998; 98EP-0200365.
XX PR 04-FEB-1998; 98EP-0200323.
XX
XX PA (COLL/) COLLEN D J.
XX PA (LEUV-) LEUVEN RES & DEV VZW.
XX
XX PI Collen DJ.
XX
XX DR WPI; 1999-508504/42.
XX
XX PT Staphylokinase derivatives with reduced immunogenicity, used for,
XX PT e.g. treatment of arterial thrombosis
XX
XX PS Claim 7; Page -: 101pp; English.
XX
XX CC The present sequence is a specifically claimed Staphylokinase SakSTAR
XX CC variant. This variant has one aminoacid that has been substituted
XX CC by another aminoacid that reduces the reactivity with monoclonal
XX CC antibodies and absorption of SakSTAR-specific antibodies from plasma
XX CC of patients treated with staphylokinase. The derivatives can also be
XX CC substituted with cysteine modified with PEG to maintain the specific
XX CC activity and significantly reduce the plasma clearance. They have
XX CC altered immunogenicity without markedly reducing the specific activity.
XX CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
XX CC potency in human plasma. The new staphylokinase derivatives are used
XX CC for treatment of arterial thrombosis, especially myocardial infarction.
XX CC The compositions can be used in human or veterinary practice.
XX CC Note: The present sequence is not shown in the specification, but is
XX CC derived from the staphylococcus aureus wild type staphylokinase sequence

```

Query Match 99.2%; Score 709; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 8.6e-70;

CC 109th Lys in SAK is replaced by Ala.
 XX Sequence 136 AA:
 SQ

Query Match 99.0%; Score 708; DB 19; Length 136;
 Best Local Similarity 98.5%; Pred. No. 1.1e-69;
 Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSEDKGKRRKGDASYEPTGPIYMNVTGVDSKGNELLSPHYVEPPIRPGTLTKEKI 60
 |||||
 Db 1 ssaftdkgkykkgddasfyepgtgplmwnvrgvgdgknellsphyvcpfpipgtlttkqkl 60
 OY 61 EYVEMALDPAKKEFRVELDPSAKIEVTYYDKNKKKEETKSPPTREKGFVPDLSSEHI 120
 |||||
 Db 61 eyvewaldataykefrveldpakievtydknkkkeetksfpftekgyfvpdlseh1 120

OY 121 KNPGFNLTKEVTEKK 136
 |||||
 Db 121 knpgfnltkvvlekk 136

RESULT 24

AAM44693
 ID AAM44693 standard; protein; 136 AA.

AC AAM44693;

DT 11-MAY-1998 (first entry)

XX Staphylokinase mutant (E73Q), useful as a thrombolytic agent.

KW Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.

OS Staphylococcus aureus.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 46 /note- "Wild-type Glu has been replaced by Gln"

XX JP10028587-A.

XX PD 03-FEB-1998.

XX PF 19-JUL-1996; 96JP-0108991.

XX PR 19-JUL-1996; 96JP-0208991.

XX PA (HONS) YAKULT HONSHA KK.

XX DR WPI; 1998-162525/15.

XX Staphylokinase mutants - useful as thrombolytic agents

XX Claim 1; Page 7; 9pp; Japanese.

CC The present sequence represents a specifically claimed staphylokinase
 CC (SAK) mutant (E73Q). The invention relates to seven staphylokinase
 CC mutants which are useful as thrombolytic agents. The mutants are: (2) a
 CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136
 CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th
 CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K113A) in which the 86th
 CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th
 CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th
 CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th
 CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the
 CC 109th Lys in SAK is replaced by Ala.

XX Sequence 136 AA:
 SQ

Query Match 99.0%; Score 708; DB 19; Length 136;

Best Local Similarity 98.5%; Pred. No. 1.1e-69;
 Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSEDKGKRRKGDASYEPTGPIYMNVTGVDSKGNELLSPHYVEPPIRPGTLTKEKI 60
 |||||
 Db 1 ssaftdkgkykkgddasfyepgtgplmwnvrgvgdgknellsphyvcpfpipgtlttkqkl 60
 OY 61 EYVEMALDPAKKEFRVELDPSAKIEVTYYDKNKKKEETKSPPTREKGFVPDLSSEHI 120
 |||||
 Db 61 eyvewaldataykefrveldpakievtydknkkkeetksfpftekgyfvpdlseh1 120
 OY 121 KNPGFNLTKEVTEKK 136
 |||||
 Db 121 knpgfnltkvvlekk 136

RESULT 25

AAV15026
 ID AAV15026 standard; protein; 136 AA.

AC AAV15026;

DT 03-NOV-1999 (first entry)

XX Staphylokinase variant SakSTAR (V132N).

KW Staphylokinase; variant; immunogenic; specifically; derivative; cysteine;

KW SakSTAR (V132N); anti-thrombotic; fibrinolytic; cardiant; veterinary;

KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;

KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.

OS Staphylococcus aureus.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 132 /note- "Wild type Val is substituted by Asn"

XX WC09940198-A2.

XX PD 12-AUG-1999.

XX PF 04-FEB-1999; 99WO-EP00748.

XX PR 06-FEB-1998; 98EP-0200365.

XX PR 04-FEB-1998; 98EP-0200323.

XX PA (COLL) COLLEN D J.

XX PA (LEUV-) LEUVEN RES & DEV VZW.

XX PI Collen DJ;

XX WPI; 1999-508504/42.

XX Staphylokinase derivatives with reduced immunogenicity, used for,

XX e.g. treatment of arterial thrombosis

XX Claim 7; Page -: 101pp; English.

CC The present sequence is a specifically claimed Staphylokinase SakSTAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SakSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the staphylococcus aureus wild type staphylokinase sequence

CC given in figure 1.
 XX Sequence 136 AA;
 SQ

Query Match 99.0%; Score 708; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 1,1e-69;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGGDASFEPTGPIYMWNVGVDSKGNELISPHYVEPIKPGTTLTKRKI 60
 |||||||
 DB 1 sssfdkgykkggdasyfepgplmynvlgvdskgnelisphyvepikpgtltlckxi 60
 |||||||
 OY 61 EYVEMALDATAYKEFRVELDPSAKIEVTYYDKNNKKETKSPITTEKGFVVDLSEHI 120
 |||||||
 DB 61 eyvewaldataykefrveldpssakievtyydknnkketksfptekgfvyvdlsehi 120
 |||||||
 OY 121 KNGCFNLITKVIIEKK 136
 |||||||
 DB 121 knpgfnliltkvilekk 136

RESULT 26

AAI15027
 ID AAI15027 standard; Protein: 136 AA.

AC AAI15027;

XX 03-NOV-1999 (first entry)

DE Staphylokinase variant SakSTAR (V132R).

XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
 KM SakSTAR (V132R); anti-thrombotic; fibrinolytic; cardiant; veterinary;
 KM Staphylococcus aureus; myocardial infarction; arterial thrombosis;
 KM monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
 XX Staphylococcus aureus.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 132 /note= "Wild type Val is substituted by Arg"

PN W09940198-A2.

PD 12-AUG-1999.

PF 04-FEB-1999; 99WO-EP00748.

PR 06-FEB-1998; 98EP-0200365.

PR 04-FEB-1998; 98EP-0200323.

XX (COLL/) COLLEN D. J.

PA (LEOV-) LEUVEN RES & DEV VZW.

PI Colleen DJ;

DR WPI; 1999-508504/42.

XX Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis

PS Claim 7; Page -; 101pp; English.

CC The present sequence is a specifically claimed Staphylokinase SakSTAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SakSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.

CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 XX

SQ Sequence 136 AA;

Query Match 99.0%; Score 708; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 1,1e-69;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGGDASFEPTGPIYMWNVGVDSKGNELISPHYVEPIKPGTTLTKRKI 60
 |||||||
 DB 1 sssfdkgykkggdasyfepgplmynvlgvdskgnelisphyvepikpgtltlckxi 60
 |||||||
 OY 61 EYVEMALDATAYKEFRVELDPSAKIEVTYYDKNNKKETKSPITTEKGFVVDLSEHI 120
 |||||||
 DB 61 eyvewaldataykefrveldpssakievtyydknnkketksfptekgfvyvdlsehi 120
 |||||||
 OY 121 KNGCFNLITKVIIEKK 136
 |||||||
 DB 121 knpgfnliltkvilekk 136

RESULT 27

AAW44694
 ID AAW44694 standard; protein: 136 AA.

AC AAW44694;

XX 11-MAY-1998 (first entry)

DE Staphylokinase mutant (D96E), useful as a thrombolytic agent.

XX Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.

XX Staphylococcus aureus.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 69 /note= "Wild-type Asp has been replaced by Glu"

PN JP10028587-A.

PD 03-FEB-1998.

PF 19-JUL-1996; 96JP-0108991.

PR 19-JUL-1996; 96JP-0208991.

XX (HONS) YAKULT HONSHA KK.

PA WPI; 1998-162525/15.

DR Staphylokinase mutants - useful as thrombolytic agents

XX Claim 1; Page 8; 9pp; Japanese.

CC The present sequence represents a specifically claimed staphylokinase
 CC (SAK) mutant (D96E). The invention relates to seven staphylokinase
 CC mutants which are useful as thrombolytic agents. The mutants are: (2) a
 CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136
 CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th
 CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K133A) in which the 56th
 CC Lys in SAK is replaced by Ser; (5) a SAK mutant (K38S) in which the 11th
 CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th
 CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th
 CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the

CC 109th lys in SAK is replaced by Ala.
 XX Sequence 136 AA;

Query Match 98.9%; Score 707; DB 19; Length 136;
 Best Local Similarity 98.5%; Pred. No. 1.4e-69;
 Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGRKRRKDDASYEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKRKI 60
 Db 1 sssfdkgrkrrkddasyfepitgplymwnvntgvdgskgnellsphyvefplkpgttltkeki 60
 OY 61 EYVEMALDPAKYKEFRVVELDPSAKIEVYTYDKKKKEETKSPFITKGFVVPDLSHI 120
 Db 61 eyvewaldataykefrvveldpksakievtydydkkkkeetksfpitkgyfvvpdlshl 120
 OY 121 KNPGENLITKRVIEKK 136
 Db 121 knpgfnlitrkviekk 136

RESULT 28

AAV15007
 ID AAV15007 standard; Protein: 136 AA.

XX AAV15007;

DT 03-NOV-1999 (first entry)

XX Staphylokinase variant SakSTAR (D82A).

XX Staphylokinase; variant; immunogenic; specificity; SakSTAR (D82A);
 KW derivative; anti-thrombotic; fibrinolytic; cardiant; veterinary;
 KM Staphylococcus aureus; myocardial infarction; arterial thrombosis; PEG;
 KW monoclonal antibody; cysteine; polyethylene glycol; plasma clearance.
 OS Staphylococcus aureus.
 XX Synthetic.

FT Key Location/Qualifiers
 FT Misc-difference 82 /note= "Wild type Asp is substituted by Ala"

XX WO9940198-A2.

PD 12-AUG-1999.

PF 04-FEB-1999; 99WO-EP00748.

PR 06-FEB-1998; 98EP-0200365.

PR 04-FEB-1998; 98EP-0200323.

XX (COLL/) COLLEN D J.

PA (LEUV-) LEUVEN RES & DEV VZW.

XX COLLEN DJ;

DR WPI; 1999-508504/42.

XX Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis

XX Claim 7; Page -; 101pp; English.

XX The present sequence is a specifically claimed Staphylokinase SakSTAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SakSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.

CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.

XX Sequence 136 AA;

Query Match 98.9%; Score 707; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 1.4e-69;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGRKRRKDDASYEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKRKI 60
 Db 1 sssfdkgrkrrkddasyfepitgplymwnvntgvdgskgnellsphyvefplkpgttltkeki 60
 OY 61 EYVEMALDPAKYKEFRVVELDPSAKIEVYTYDKKKKEETKSPFITKGFVVPDLSHI 120
 Db 61 eyvewaldataykefrvveldpksakievtydydkkkkeetksfpitkgyfvvpdlshl 120
 OY 121 KNPGENLITKRVIEKK 136
 Db 121 knpgfnlitrkviekk 136

RESULT 29

AAV15018
 ID AAV15018 standard; Protein: 136 AA.

XX AAV15018;

DT 03-NOV-1999 (first entry)

XX Staphylokinase variant SakSTAR (G36R).

XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
 KW SakSTAR (G36R); anti-thrombotic; fibrinolytic; cardiant; veterinary;
 KM Staphylococcus aureus; myocardial infarction; arterial thrombosis;
 KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
 OS Staphylococcus aureus.
 XX Synthetic.

FT Key Location/Qualifiers
 FT Misc-difference 36 /note= "Wild type Gly is substituted by Arg"

XX WO9940198-A2.

PD 12-AUG-1999.

PF 04-FEB-1999; 99WO-EP00748.

PR 06-FEB-1998; 98EP-0200365.

PR 04-FEB-1998; 98EP-0200323.

XX (COLL/) COLLEN D J.

PA (LEUV-) LEUVEN RES & DEV VZW.

XX COLLEN DJ;

DR WPI; 1999-508504/42.

XX Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis

XX Claim 7; Page -; 101pp; English.

XX The present sequence is a specifically claimed staphylokinase SakSTAR
 CC variant. This variant has one aminoacid that has been substituted

CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SakSTAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.

CC Sequence 136 AA;

Query Match 98.9%; Score 707; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 1,4e-69;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSPDKGKYYKKGDADSYFEPTGPIYLMVNVGVDSKGNELLSPHYVEPIKPGTTLTKEXI 60
 |||
 Db 1 sssfkgykkgddasyleptgpylmvntgvdsaknellspkyvefpikpgtltltxeki 60
 |||
 OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVDLSEHI 120
 |||
 Db 61 eyvewaldataykefrvveldpesakievtyydknkkkeetksfptekgfivvpdlsehl 120
 |||
 OY 121 KNGFNLITKRVIEKK 136
 |||
 Db 121 knpgfnliltkvvlekk 136
 |||

RESULT 30

AA15019

ID AAY15019 standard; Protein: 136 AA.

AC AAY15019;

XX 03-NOV-1999 (first entry)

DE Staphylokinase variant SakSTAR (H43R).

XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;

KW SakSTAR (H43R); anti-thrombotic; fibrinolytic; cardiant; veterinary;

KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;

KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.

XX Staphylococcus aureus.

OS Synthetic.

XX Key

FT Misc-difference 43 /note- "Wild type His is substituted by Arg"

XX MO9940198-A2.

XX 12-AUG-1999.

XX 04-FEB-1999; 99WO-EP00748.

XX 06-FEB-1998; 98EP-0200365.

XX 04-FEB-1998; 98EP-0200323.

XX (COLL/) COLLEN D. J.

XX (LEUV-) LEUVEN RES & DEV VZW.

XX Collen DJ;

XX WPI; 1999-508504/42.

XX Staphylokinase derivatives with reduced immunogenicity, used for,

PT e.g. treatment of arterial thrombosis

XX Claim 7; Page -: 101pp; English.

XX The present sequence is a specifically claimed Staphylokinase SakSTAR

XX variant. This variant has one aminoacid that has been substituted

XX by another aminoacid that reduces the reactivity with monoclonal

XX antibodies and absorption of SakSTAR-specific antibodies from plasma

XX of patients treated with staphylokinase. The derivatives can also be

XX substituted with cysteine modified with PEG to maintain the specific

XX activity and significantly reduce the plasma clearance. They have

XX altered immunogenicity without markedly reducing the specific activity.

XX This sequence has anti-thrombotic, cardiant activity and a fibrinolytic

XX potency in human plasma. The new staphylokinase derivatives are used

XX for treatment of arterial thrombosis, especially myocardial infarction.

XX The compositions can be used in human or veterinary practice.

XX Note: The present sequence is not shown in the specification, but is

XX derived from the Staphylococcus aureus wild type staphylokinase sequence

XX given in figure 1.

CC Sequence 136 AA;

Query Match 98.9%; Score 707; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 1,4e-69;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSPDKGKYYKKGDADSYFEPTGPIYLMVNVGVDSKGNELLSPHYVEPIKPGTTLTKEXI 60
 |||
 Db 1 sssfkgykkgddasyleptgpylmvntgvdsaknellspkyvefpikpgtltltxeki 60
 |||
 OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVDLSEHI 120
 |||
 Db 61 eyvewaldataykefrvveldpesakievtyydknkkkeetksfptekgfivvpdlsehl 120
 |||
 OY 121 KNGFNLITKRVIEKK 136
 |||
 Db 121 knpgfnliltkvvlekk 136
 |||

RESULT 31

AA15011

ID AAY15011 standard; Protein: 136 AA.

AC AAY15011;

XX 03-NOV-1999 (first entry)

DE Staphylokinase variant SakSTAR (D82A).

XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;

KW SakSTAR (D82A); anti-thrombotic; fibrinolytic; cardiant; veterinary;

KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;

KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.

XX Staphylococcus aureus.

OS Synthetic.

XX Key

FT Misc-difference 82 /note- "Wild type Asp is substituted by Ala"

XX MO9940198-A2.

XX 12-AUG-1999.

XX 04-FEB-1999; 99WO-EP00748.

XX 06-FEB-1998; 98EP-0200365.

XX 04-FEB-1998; 98EP-0200323.

XX (COLL/) COLLEN D. J.

XX (LEUV-) LEUVEN RES & DEV VZW.

XX Collen DJ;
 PI
 XX WPI: 1999-508504/42.
 DR
 XX
 PT Staphylokinase derivatives with reduced immunogenicity, used for,
 XX e.g. treatment of arterial thrombosis
 PS
 XX Claim 7: Page -: 101pp; English.
 CC
 CC The present sequence is a specifically claimed staphylokinase SaksNAR
 CC variant. This variant has one aminoacid that has been substituted
 CC by another aminoacid that reduces the reactivity with monoclonal
 CC antibodies and absorption of SaksNAR-specific antibodies from plasma
 CC of patients treated with staphylokinase. The derivatives can also be
 CC substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used
 CC for treatment of arterial thrombosis, especially myocardial infarction.
 CC The compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 CC
 SQ Sequence 136 AA;

Query Match 98.9%; Score 707; DB 20; Length 136;
 Best Local Similarity 99.3%; Pred. No. 1.4e-69;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSSFDKGRKKGDASFEPTGPTLMVNTGVDSKGNELSPHYVEFPKPGTTLTKEXI 60
 DB 1 sssfdkgrkkgddasfepptgpylmvntgvdskgnellsphyvefpkpgtltltek1 60
 QY 61 EYVEWALDPAKYEFRRVELDPSAKIEVTYYDNKKKEETKSFPITTEKGFVPDLSEHI 120
 DB 61 eyvewaldataykefrvelapsakievtydknkkkeetksfpitkxgfvpdlsehl 120
 QY 121 KNPGFNLTIKVIEKK 136
 DB 121 knpgfnltikvlekk 136

RESULT 32
 AAM44689
 ID AAM44689 standard; protein; 136 AA.
 AC AAM44689;

XX 11-MAY-1998 (first entry)
 DE Staphylokinase mutant (D41N), useful as a thrombolytic agent.
 KW Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.
 OS Synthetic.
 OS
 XX
 FT Key Location/Qualifiers
 FT Modified-site 14 /note= "Wild-type Asp has been replaced by Asn"

XX JP10028587-A.
 XX
 XX 03-FEB-1998.
 XX
 XX 19-JUL-1996; 96JP-0208991.
 XX
 XX 19-JUL-1996; 96JP-0208991.
 XX

PA (HONS) YAKULT HONSHA KK.
 XX
 XX WPI: 1998-162525/15.
 DR
 XX
 PT Staphylokinase mutants - useful as thrombolytic agents
 XX
 PS Claim 1; Pages 5-6; 9pp; Japanese.
 PS
 XX

CC The present sequence represents a specifically claimed staphylokinase
 CC (SAK) mutant (D41N). The invention relates to seven staphylokinase
 CC mutants which are useful as thrombolytic agents. The mutants are: (2) a
 CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136
 CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th
 CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K13A) in which the 86th
 CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th
 CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th
 CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th
 CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the
 CC 109th Lys in SAK is replaced by Ala.
 CC
 SQ Sequence 136 AA;

Query Match 98.7%; Score 706; DB 19; Length 136;
 Best Local Similarity 98.5%; Pred. No. 1.8e-69;
 Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSSFDKGRKKGDASFEPTGPTLMVNTGVDSKGNELSPHYVEFPKPGTTLTKEXI 60
 DB 1 sssfdkgrkkgddasfepptgpylmvntgvdskgnellsphyvefpkpgtltltek1 60
 QY 61 EYVEWALDPAKYEFRRVELDPSAKIEVTYYDNKKKEETKSFPITTEKGFVPDLSEHI 120
 DB 61 eyvewaldataykefrvelapsakievtydknkkkeetksfpitkxgfvpdlsehl 120
 QY 121 KNPGFNLTIKVIEKK 136
 DB 121 knpgfnltikvlekk 136

RESULT 33
 AAM44692
 ID AAM44692 standard; protein; 136 AA.
 AC AAM44692;

XX 11-MAY-1998 (first entry)
 DE Staphylokinase mutant (K38S), useful as a thrombolytic agent.
 KW Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.
 OS Synthetic.
 OS
 XX
 FT Key Location/Qualifiers
 FT Modified-site 11 /note= "Wild-type Lys has been replaced by Ser"

XX JP10028587-A.
 XX
 XX 03-FEB-1998.
 XX
 XX 19-JUL-1996; 96JP-0108991.
 XX
 XX 19-JUL-1996; 96JP-0208991.
 XX
 XX (HONS) YAKULT HONSHA KK.
 XX
 XX WPI: 1998-162525/15.
 XX
 XX Staphylokinase mutants - useful as thrombolytic agents
 XX

PS Claim 1; Page 7; 9pp; Japanese.

XX The present sequence represents a specifically claimed staphylokinase
CC (SAK) mutant (K138S). The invention relates to seven staphylokinase
CC mutants which are useful as thrombolytic agents. The mutants are: (2) a
CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136
CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th
CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K113A) in which the 86th
CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th
CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th
CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th
CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the
CC 109th Lys in SAK is replaced by Ala.

XX Sequence 136 AA;
SQ

Query Match 98.7%; Score 706; DB 19; Length 136;
Best Local Similarity 98.5%; Pred. No. 1.8e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSFDDKGRYKKGGDASYEPPTGYLMVNTGVDSKGNELLSPHYVEFPPIKPGTTLTKEXI 60
|||
Db 1 sssfdkgkyksgddasyfepptgylmwnvntgvdgkgnellsphyvefpikpgtltltxekl 60

QY 61 EYVEMALDQATAYKEFRVVELDPSAKIEVTYDKNKKKEETKSFPITKEGFVVPDLSEHI 120
|||
Db 61 eyvewaldataykefrvveldpakievtydknkkkeetksfptekgfvpdlsehl 120

QY 121 KNPGFNILTKVVIKK 136
|||
Db 121 knpgfniltkvvielk 136

RESULT 34
AA15028
ID AA15028 standard; Protein: 136 AA.
XX
AC AA15028;
XX
DT 03-NOV-1999 (first entry)
XX
DE Staphylokinase variant SakSTAR (K130T, K135R).
XX
KW Staphylokinase; immunogenic; specificity; SakSTAR (K130T, K135R);
KW variant; derivative; anti-thrombotic; fibrinolytic; cardiant; veterinary;
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis; PEG;
KW monoclonal antibody; cysteine; polyethylene glycol; plasma clearance.
XX
OS Staphylococcus aureus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT MISC-difference 130 /note- "Wild type Lys is substituted by Thr"
FT MISC-difference 135 /note- "Wild type Lys is substituted by Arg"
FT
XX
FN MO9940198-A2.
XX
PD 12-AUG-1999.
XX
PF 04-FEB-1999; 99WO-EP00748.
XX
PR 06-FEB-1998; 98EP-0200365.
PR 04-FEB-1998; 98EP-0200323.
XX
PA (COLL/) COLLEN D J.
PA (LEUV-) LEUVEN RES & DEV VZW.
XX
PI Collen DJ;
XX
DR WPI: 1999-508504/42.

XX Staphylokinase derivatives with reduced immunogenicity, used for,
PT e.g. treatment of arterial thrombosis
XX
PS Claim 7; Page -; 10lpp; English.

XX The present sequence is a specifically claimed Staphylokinase SakSTAR
CC variant. This variant has more than one aminoacid that has been
CC substituted by another aminoacid that reduces the reactivity with
CC monoclonal antibodies and absorption of SakSTAR-specific antibodies from
CC plasma of patients treated with staphylokinase. The derivatives can also
CC be substituted with cysteine modified with PEG to maintain the specific
CC activity and significantly reduce the plasma clearance. They have
CC altered immunogenicity without markedly reducing the specific activity.
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
CC potency in human plasma. The new staphylokinase derivatives are used for
CC treatment of arterial thrombosis, especially myocardial infarction. The
CC compositions can be used in human or veterinary practice.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the Staphylococcus aureus wild type staphylokinase sequence
CC given in figure 1.
XX
SQ Sequence 136 AA;

Query Match 98.7%; Score 706; DB 20; Length 136;
Best Local Similarity 98.5%; Pred. No. 1.8e-69;
Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSFDDKGRYKKGGDASYEPPTGYLMVNTGVDSKGNELLSPHYVEFPPIKPGTTLTKEXI 60
|||
Db 1 sssfdkgkyksgddasyfepptgylmwnvntgvdgkgnellsphyvefpikpgtltltxekl 60

QY 61 EYVEMALDQATAYKEFRVVELDPSAKIEVTYDKNKKKEETKSFPITKEGFVVPDLSEHI 120
|||
Db 61 eyvewaldataykefrvveldpakievtydknkkkeetksfptekgfvpdlsehl 120

QY 121 KNPGFNILTKVVIKK 136
|||
Db 121 knpgfniltkvvielk 136

RESULT 35
AA15014
ID AA15014 standard; Protein: 136 AA.
XX
AC AA15014;
XX
DT 03-NOV-1999 (first entry)
XX
DE Staphylokinase variant SakSTAR (Y92A).
XX
KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KW SakSTAR (Y92A); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
XX
OS Staphylococcus aureus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT MISC-difference 92 /note- "Wild type Tyr is substituted by Ala"
FT
XX
FN MO9940198-A2.
XX
PD 12-AUG-1999.
XX
PF 04-FEB-1999; 99WO-EP00748.
XX
PR 06-FEB-1998; 98EP-0200365.
PR 04-FEB-1998; 98EP-0200323.
XX

PA	(COLL/) COLLEN D J.
PA	(LEUV-) LEUVEN RES & DEV VZW.
XX	
PI	Collen DJ:
XX	
DR	WPI: 1999-508504/42.
XX	
PR	Staphylokinase derivatives with reduced immunogenicity, used for,
PT	e.g. treatment of arterial thrombosis
XX	
PS	Claim 7; Page -: 101pp; English.
XX	
CC	The present sequence is a specifically claimed Staphylokinase SakSPAR
CC	variant. This variant has one aminoacid that has been substituted
CC	by another aminoacid that reduces the reactivity with monoclonal
CC	antibodies and absorption of SakSPAR-specific antibodies from plasma
CC	of patients treated with staphylokinase. The derivatives can also be
CC	substituted with cysteine modified with PEG to maintain the specific
CC	activity and significantly reduce the plasma clearance. They have
CC	altered immunogenicity without markedly reducing the specific activity.
CC	This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
CC	potency in human plasma. The new staphylokinase derivatives are used
CC	for treatment of arterial thrombosis, especially myocardial infarction.
CC	The compositions can be used in human or veterinary practice.
CC	Note: The present sequence is not shown in the specification, but is
CC	derived from the Staphylococcus aureus wild type staphylokinase sequence
CC	given in figure 1.
XX	
SQ	Sequence 136 AA;
	Query Match 98.7%; Score 706; DB 20; Length 136;
	Best Local Similarity 99.3%; Pred.No.1.8e-69;
	Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
OY	1 SSSFDKGRKKKGDDASYEPTGPRYLMMVNTGVDSKGNELLSPHYVEFPKPQTLLKEKI 60
Db	1 SSSFdkgrykkkdasyfepigpylmmvntgydskgnellspkyvefpkpgtlttketl 60
OY	61 EYYVMALDATNYKFFRYVELDPSAKIEVTYYDKNKKKEETKSFTTEKGFVPDLSEHI 120
Db	61 eyyvmaldatataykfrryveldpsakievtyadknkkkeetsftltekgfvvpdlsehl 120
OY	121 KNPGNLITKVYIEKK 136
Db	121 knpgnilitkvviekk 136
RESULT 36	
AAB01297	
ID AAB01297	standard; protein: 136 AA.
XX	
AC AAB01297:	
XX	
DT 25-SEP-2000	(first entry)
XX	
DE Wild type staphylokinase from Staphylococcus aureus.	
XX	
KW Immunogenicity; immunogen; T cell epitope; T-lymphocyte; drug;	
KW vaccine; carrier; diagnosis; therapy; complement; C3; C5; CVF;	
KW plasminogen; staphylokinase; fibrin; blood clot; thrombolysis;	
KW plasmin; myocardial infarction.	
XX	
OS Staphylococcus aureus.	
XX	
PN WO200034317-A2.	
XX	
PD 15-JUN-2000.	
XX	
PF 08-DEC-1999;	99WO-GB04119.
XX	
XX 08-DEC-1998;	98GB-0026925.
PR 02-FEB-1999;	99GB-0002139.
PR	

PA	(BIOV-) BIOVATION LTD.
XX	
XX	Carr FU, Adair FS, Hamilton AA, Carter G;
XX	
XX	WPI; 2000-423372/36.
DR	
XX	
XX	Rendering a protein non-immunogenic or less immunogenic useful in
PT	medicine and in diagnostics involves determining the amino acid
PT	sequence of the protein, identifying and modifying potential epitopes
XX	
PS	Example 3; Fig 7; 42pp; English.
XX	
CC	Proteins or their fragments can be rendered non-immunogenic or less
CC	immunogenic by identifying one or more potential T cell epitopes and
CC	modifying the sequence to eliminate at least one of these T cell
CC	epitopes to reduce the immunogenicity of the protein when exposed to
CC	the immune system of another organism. The method can be used for
CC	producing a protein with an enzymatic activity which has a beneficial
CC	therapeutic effect, a protein used to convert inactive drugs to its
CC	active form within a living organism, a vaccine, a protein used as a
CC	carrier of other molecule or a protein which binds to other molecules
CC	within or introduced within the living organism in order to alter the
CC	bio distribution of other molecules, such that the protein does not
CC	stimulate immune response in the living organism. The less immunogenic
CC	protein is useful in medicine, diagnosis and in manufacture of a
CC	therapeutic or a diagnostic agent. Staphylokinase is produced by
CC	Staphylococcus aureus. The protein holds considerable promise for
CC	thrombolytic therapy. Clinical use in humans has been limited due to
CC	the native protein being immunogenic. Availability of a
CC	non-immunogenic staphylokinase would have considerable importance as
CC	a potential agent for thrombolytic therapy. This is the sequence
CC	of the wild type staphylokinase. The altered staphylokinase sequence
CC	is given in AAB01298. See GENSEQ records AAB01289-B01302.
XX	
SQ	Sequence 136 AA;
Query Match	98.7%; Score 706; DB 21; Length 136;
Best Local Similarity	99.3%; Pred. 1.8e-69;
Matches 135; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 SSSPFGKKKKDDDDSYEPETPYLMVNVWGDSKGNELLSPHYVEPIKGTTLTKRKI 60
DB	1 SSSfKkykYkKdDdSyfepcgyLMvNvGdSgNellSPHYVEPIKpGTLTKRKi 60
QY	61 EYXVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSPFITRKGFVVPDLSHI 120
DB	61 EYyVemAlDatAyKefrVvElDpSakIeVtYdKnKkKeeTksfPITkGfVvPdlShI 120
QY	121 KNPGENLITKVVIEKK 136
DB	121 kNpGnElItKvVIEkK 136
RESULT 37	
AAMW44691	
ID	AAMW44691 standard; protein; 136 AA.
XX	
XX	AAMW44691;
XX	
DT	11-MAY-1998 (first entry)
XX	
DE	Staphylokinase mutant (K113A), useful as a thrombolytic agent.
XX	
KW	Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.
OS	Staphylococcus aureus.
OS	Synthetic.
XX	
XX	
FIH	Key Location/Qualifiers
FT	Modified-site 86
FT	/note= "wild-type Lys has been replaced by Ala"


```

XX  JP10028587-A.
PN
XX
XX  03-FEB-1998.
PD
XX
XX  19-JUL-1996; 96JP-0108991.
PE
XX
XX  19-JUL-1996; 96JP-0208991.
PR
XX
XX  (HONS ) YAKULT HONSHA KK.
PA
XX
XX  WPI; 1998-162525/15.
DR
XX
XX  Staphylokinase mutants - useful as thrombolytic agents
PT
XX
XX  Claim 1; Pages 6-7; 9pp; Japanese.
PS
XX
CC  The present sequence represents a specifically claimed staphylokinase
CC  (SAK) mutant (K136A). The invention relates to seven staphylokinase
CC  mutants which are useful as thrombolytic agents. The mutants are: (2) a
CC  SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136
CC  amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th
CC  Glu in SAK is replaced by Gln; (4) a SAK mutant (K136A) in which the 86th
CC  Lys in SAK is replaced by Ser; (5) a SAK mutant (K38S) in which the 11th
CC  Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th
CC  Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th
CC  Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the
CC  109th Lys in SAK is replaced by Ala.
CC
XX
XX  Sequence 136 AA;
SO

Query Match          98.6%; Score 705; DB 19; Length 136;
Best Local Similarity 98.5%; Pred. No. 2,4e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY  1 SSSPDKGKTKKGDASTYEPGTYLWVNTGVDSKGNELSPHYVEPIKPGTTLTKRKI 60
    |||||||
DB  1 sssldkgykkgdastyleptgpylmvntvgdgkgnelssphyvefplkpgtltltek1 60

OY  61 EYVEMALDATAKKEFRVVELDPSAKIEVYDYDNKKKKEETSPITEKGFVPDISEHI 120
    |||||||
DB  61 eyvewaldataykefrvveldpasaalevttydknkkkeetskfpitcektfvydpdiseh1 120

OY  121 KNPGNLTITKVIEKK 136
    |||||||
DB  121 knpgfnliltkvilekk 136

RESULT 38
AAW44695
ID  AAW44695 standard; protein; 136 AA.
XX
XX  AAW44695;
AC
XX
XX  11-MAY-1998 (first entry)
DT
XX
XX  Staphylokinase mutant (K136A), useful as a thrombolytic agent.
DE
XX
XX  Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.
KW
XX
XX  Staphylococcus aureus.
OS
XX
XX  Synthetic.
OS
XX
XX  Key Location/Qualifiers
FH  Modified-site 109
FT  /note="Wild-type Lys has been replaced by Ala"
XX
XX  JP10028587-A.
XX
XX  03-FEB-1998.
XX
XX  19-JUL-1996; 96JP-0108991.
XX

```

```

XX  19-JUL-1996; 96JP-0208991.
XX
XX  (HONS ) YAKULT HONSHA KK.
PA
XX
XX  WPI; 1998-162525/15.
DR
XX
XX  Staphylokinase mutants - useful as thrombolytic agents
PT
XX
XX  Claim 1; Page 8; 9pp; Japanese.
PS
XX
CC  The present sequence represents a specifically claimed staphylokinase
CC  (SAK) mutant (K136A). The invention relates to seven staphylokinase
CC  mutants which are useful as thrombolytic agents. The mutants are: (2) a
CC  SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136
CC  amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th
CC  Glu in SAK is replaced by Gln; (4) a SAK mutant (K136A) in which the 86th
CC  Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th
CC  Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th
CC  Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th
CC  Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the
CC  109th Lys in SAK is replaced by Ala.
CC
XX
XX  Sequence 136 AA;
SO

Query Match          98.6%; Score 705; DB 19; Length 136;
Best Local Similarity 98.5%; Pred. No. 2,4e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY  1 SSSPDKGKTKKGDASTYEPGTYLWVNTGVDSKGNELSPHYVEPIKPGTTLTKRKI 60
    |||||||
DB  1 sssldkgykkgdastyleptgpylmvntvgdgkgnelssphyvefplkpgtltltek1 60

OY  61 EYVEMALDATAKKEFRVVELDPSAKIEVYDYDNKKKKEETSPITEKGFVPDISEHI 120
    |||||||
DB  61 eyvewaldataykefrvveldpasaalevttydknkkkeetskfpitceagfvydpdiseh1 120

OY  121 KNPGNLTITKVIEKK 136
    |||||||
DB  121 knpgfnliltkvilekk 136

RESULT 39
AAAY15041
ID  AAAY15041 standard; Protein; 136 AA.
XX
XX  AAAY15041;
AC
XX
XX  03-NOV-1999 (first entry)
DT
XX
XX  Staphylokinase variant SakSTAR (K74Q, K130E, K135R).
DE
XX
XX  Staphylokinase; immunogenic; specificity; SakSTAR (K74Q, K130E, K135R);
KW  variant; derivative; anti-thrombotic; fibrinolytic; cardiant; veterinary;
KW  Staphylococcus aureus; myocardial infarction; arterial thrombosis; PEG;
KW  monoclonal antibody; cysteine; polyethylene glycol; plasma clearance.
XX
XX  Staphylococcus aureus.
OS
XX
XX  Synthetic.
OS
XX
XX  Key Location/Qualifiers
FH  Misc-difference 74
FT  /note="Wild type Lys is substituted by Gln"
XX
XX  Misc-difference 130
FT  /note="Wild type Lys is substituted by Glu"
XX
XX  Misc-difference 135
FT  /note="Wild type Lys is substituted by Arg"
XX
XX  WO940198-A2.
XX
XX  12-AUG-1999.
XX

```

PF 04-FEB-1999; 99MO-EP00748.
 XX
 PR 06-FEB-1998; 98EP-0200365.
 XX 04-FEB-1998; 98EP-0200323.
 XX
 PA (COLL/) COLLEN D J.
 XX (LEUV-) LEUVEN RES & DEV VZW.
 XX
 PI Collen DJ;
 XX
 DR WPI: 1999-508504/42.
 XX
 PT Staphylokinase derivatives with reduced immunogenicity, used for,
 XX e.g. treatment of arterial thrombosis
 PS
 XX Claim 7; Page -: 101pp; English.
 CC The present sequence is a specifically claimed Staphylokinase SakSTAR
 CC variant. This variant has more than one aminoacid that has been
 CC substituted by another aminoacid that reduces the reactivity with
 CC monoclonal antibodies and absorption of SakSTAR-specific antibodies from
 CC plasma of patients treated with staphylokinase. The derivatives can also
 CC be substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used for
 CC treatment of arterial thrombosis, especially myocardial infarction. The
 CC compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 XX
 XX Sequence 136 AA;
 SQ
 Query Match 98.5%; Score 704; DB 20; Length 136;
 Best Local Similarity 97.8%; Pred. No. 3e-69;
 Matches 133; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSSPDKGKYYKGGDASYPEPTGPLYMNVNGVDSKGNELLSPHYVEFPPIKPGTTLTKEXI 60
 DB 1 sssfdkgkykkgddasyfepitgpylmvntvgvdsksnellspkphvefpikpgttltkexi 60
 QY 61 EYVEWALDATATKERRVVELDPSAKIEVYYDKNKKKEETSPITEKGFVVPDLSEHI 120
 DB 61 eyvewaldatayqefrvelldpsakievlyydknkkkeetksfptekgfvpdlsehl 120
 QY 121 KNPGFNLTITKXVIEKK 136
 DB 121 knpgfnltitkvierk 136
 RESULT 40
 ID AAY15020 standard; Protein; 136 AA.
 XX AAY15020;
 AC
 XX 03-NOV-1999 (first entry)
 DT
 XX Staphylokinase variant SakSTAR (G36R, K74R).
 DE
 XX Staphylokinase: variant; immunogenic; specificity; derivative; cysteine;
 KW SakSTAR (G36R, K74R); anti-thrombotic; fibrinolytic; cardiant;
 KM Staphylococcus aureus; myocardial infarction; arterial thrombosis;
 XX monoclinal antibody; polyethylene glycol; PEG; plasma clearance.
 OS Staphylococcus aureus.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 36

FT /note- "Wild type Gly is substituted by Arg"
 FT Misc-difference 74
 FT /note- "Wild type Lys is substituted by Arg"
 XX
 XX W09940198-A2.
 XX
 XX 12-AUG-1999.
 PD
 XX
 XX 04-FEB-1999; 99MO-EP00748.
 XX
 XX 06-FEB-1998; 98EP-0200365.
 XX 04-FEB-1998; 98EP-0200323.
 XX
 XX (COLL/) COLLEN D J.
 XX (LEUV-) LEUVEN RES & DEV VZW.
 XX
 XX Collen DJ;
 PI
 XX WPI: 1999-508504/42.
 DR
 XX Staphylokinase derivatives with reduced immunogenicity, used for,
 XX e.g. treatment of arterial thrombosis
 PT
 XX Claim 7; Page -: 101pp; English.
 PS
 XX The present sequence is a specifically claimed Staphylokinase SakSTAR
 XX variant. This variant has more than one aminoacid that has been
 XX substituted by another aminoacid that reduces the reactivity with
 XX monoclonal antibodies and absorption of SakSTAR-specific antibodies from
 XX plasma of patients treated with staphylokinase. The derivatives can also
 XX be substituted with cysteine modified with PEG to maintain the specific
 XX activity and significantly reduce the plasma clearance. They have
 XX altered immunogenicity without markedly reducing the specific activity.
 XX This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 XX potency in human plasma. The new staphylokinase derivatives are used for
 XX treatment of arterial thrombosis, especially myocardial infarction. The
 XX compositions can be used in human or veterinary practice.
 XX Note: The present sequence is not shown in the specification, but is
 XX derived from the Staphylococcus aureus wild type staphylokinase sequence
 XX given in figure 1.
 XX
 XX Sequence 136 AA;
 SQ
 Query Match 98.5%; Score 704; DB 20; Length 136;
 Best Local Similarity 97.8%; Pred. No. 3e-69;
 Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SSSPDKGKYYKGGDASYPEPTGPLYMNVNGVDSKGNELLSPHYVEFPPIKPGTTLTKEXI 60
 DB 1 sssfdkgkykkgddasyfepitgpylmvntvgvdsksnellspkphvefpikpgttltkexi 60
 QY 61 EYVEWALDATATKERRVVELDPSAKIEVYYDKNKKKEETSPITEKGFVVPDLSEHI 120
 DB 61 eyvewaldatayqefrvelldpsakievlyydknkkkeetksfptekgfvpdlsehl 120
 QY 121 KNPGFNLTITKXVIEKK 136
 DB 121 knpgfnltitkvierk 136
 RESULT 41
 ID AAY15012 standard; Protein; 136 AA.
 XX AAY15012;
 AC
 XX 03-NOV-1999 (first entry)
 DT
 XX Staphylokinase variant SakSTAR (D82A, S84A).
 DE
 XX Staphylokinase: variant; immunogenic; specificity; SakSTAR (D82A, S84A);
 KW derivative; anti-thrombotic; fibrinolytic; cardiant; veterinary;

KM Staphylococcus aureus; myocardial infarction; arterial thrombosis; PEG;
 KW monoclonal antibody; cysteine; polyethylene glycol; plasma clearance.
 OS Staphylococcus aureus.
 OS Synthetic.
 XX Key
 FH Location/Qualifiers
 FT Misc-difference 82
 FT /note= "Wild type Asp 1s substituted by Ala"
 FT Misc-difference 84
 FT /note= "Wild type Ser is substituted by Ala"
 FT
 XX WO9940198-A2.
 PN 12-AUG-1999.
 PD
 XX
 PF 04-FEB-1999; 99WO-EP00748.
 XX
 PR 06-FEB-1998; 98EP-0200365.
 PR 04-FEB-1998; 98EP-0200323.
 XX
 PA (COLL/) COLLEN D. J.
 PA (LEUV-) LEUVEN RES & DEV VZW.
 XX
 PI Collen DJ;
 XX
 DR WPI; 1999-508504/42.
 XX
 PT Staphylokinase derivatives with reduced immunogenicity, used for,
 PT e.g. treatment of arterial thrombosis
 PS
 PS Claim 7; Page -: 101pp; English.
 XX
 CC The present sequence is a specifically claimed Staphylokinase SakSTAR
 CC variant. This variant has more than one aminoacid that has been
 CC substituted by another aminoacid that reduces the reactivity with
 CC monoclonal antibodies and absorption of SakSTAR-specific antibodies from
 CC plasma of patients treated with staphylokinase. The derivatives can also
 CC be substituted with cysteine modified with PEG to maintain the specific
 CC activity and significantly reduce the plasma clearance. They have
 CC altered immunogenicity without markedly reducing the specific activity.
 CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
 CC potency in human plasma. The new staphylokinase derivatives are used for
 CC treatment of arterial thrombosis, especially myocardial infarction. The
 CC compositions can be used in human or veterinary practice.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the Staphylococcus aureus wild type staphylokinase sequence
 CC given in figure 1.
 CC
 CC
 SQ Sequence 136 AA;
 XX
 XX
 Query Match 98.5%; Score 704; DB 20; Length 136;
 Best Local Similarity 98.5%; Pred. No. 3e-69;
 Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

AC AAM03079;
 XX
 DT 19-FEB-1997 (first entry)
 XX
 XX Staphylokinase derivative M10 with altered charge cluster 10.
 DE Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;
 KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.
 XX
 OS Staphylococcus aureus.
 OS Synthetic.
 XX Key
 FH Location/Qualifiers
 FT Region 86..88
 FT /label= altered M10 cluster
 FT /note= "wild-type cluster 10 sequence Lys-11e-Glu
 FT has been replaced by Ala-11e-Ala to reduce
 FT immunogenicity"
 FT
 XX EP721982-A1.
 PN 17-JUL-1996.
 PD
 XX
 XX 06-JAN-1995; 95EP-0200023.
 XX
 PR 06-JAN-1995; 95EP-0200023.
 XX
 PA (COLL/) COLLEN D. J.
 PA (LEUV-) LEUVEN RES & DEV VZW.
 XX
 PI Collen D;
 XX
 DR WPI; 1996-322832/33.
 XX
 PT New staphylokinase mutants with reduced immunogenicity - useful for
 PT treating arterial thrombosis
 PS
 PS Example 2; Page -: 21pp; English.
 XX
 CC Staphylokinase derivatives showing a reduced immunogenicity as
 CC compared to wild-type staphylokinase are claimed. The derivatives
 CC are useful as thrombolytic agents to treat arterial thrombosis and
 CC are pref. produced by eliminating at least one of the epitopes
 CC indicated in the features table. The epitopes are destroyed by
 CC replacing one or more amino acid residues in a charge cluster by an
 CC Ala residue. Mutations are introduced using site-directed
 CC mutagenesis on wild-type staphylokinase genomic DNA from the S. aureus
 CC lysogenic strain SakSTAR. The present sequence is a specific
 CC example of a staphylokinase derivative according to the invention;
 CC the sequence does not appear in the specification but can be generated
 CC from the wild-type staphylokinase sequence using the description given
 CC in Table 1 of the disclosure.
 CC
 CC
 SQ Sequence 136 AA;
 XX
 XX
 Query Match 98.3%; Score 703; DB 17; Length 136;
 Best Local Similarity 98.5%; Pred. No. 3.9e-69;
 Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
RESULT 43
AAW03078 standard; protein; 136 AA.
XX
AC AAW03078;
XX
DT 19-FEB-1997 (first entry)
XX
DE Staphylokinase derivative M6 with altered charge cluster 6.
XX
KM Staphylokinase; mutant; mutein; variant; immunogenicity: decrease;
XX derivative; SakSTAR; arterial thrombosis; thrombolytic agent.
OS Staphylococcus aureus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT 61..65
FT Region /label= altered_M6_cluster
FT /note= "wild-type cluster 6 sequence
FT Glu-Tyr-Tyr-Val-Glu has been replaced by
FT Ala-Tyr-Tyr-Val-Ala to reduce immunogenicity"
XX
PN EP721982-A1.
XX
PD 17-JUL-1996.
XX
PF 06-JAN-1995; 95EP-0200023.
XX
PR 06-JAN-1995; 95EP-0200023.
XX
PA (COLL/) COLLEN D J.
XX (LEUV-) LEUVEN RES & DEV VZW.
XX
PI Collen D;
XX
DR WPI; 1996-322832/33.
XX
PT New staphylokinase mutants with reduced immunogenicity - useful for
XX treating arterial thrombosis
XX
PS Example 2; Page -: 21pp; English.
XX
CC Staphylokinase derivatives showing a reduced immunogenicity as
CC compared to wild-type staphylokinase are claimed. The derivatives
CC are useful as thrombolytic agents to treat arterial thrombosis and
CC are pref. produced by eliminating at least one of the epitopes
CC indicated in the features table. The epitopes are destroyed by
CC replacing one or more amino acid residues in a charge cluster by an
CC Ala residue. Mutations are introduced using site-directed
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus
CC lysogenic strain SakSTAR. The present sequence is a specific
CC example of a staphylokinase derivative according to the invention;
CC the sequence does not appear in the specification but can be generated
CC from the wild-type staphylokinase sequence using the description given
CC in Table 1 of the disclosure.
XX
SQ Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;
Best Local Similarity 98.5%; Pred. No. 3, 9e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
|||||
Db 121 knpgfnlitrkvlekk 136

RESULT 44
AAW03082 standard; protein; 136 AA.
XX
AC AAW03082;
XX
DT 19-FEB-1997 (first entry)
XX
DE Staphylokinase derivative M15 with altered charge cluster 15.
XX
KM Staphylokinase; mutant; mutein; variant; immunogenicity: decrease;
XX derivative; SakSTAR; arterial thrombosis; thrombolytic agent.
OS Staphylococcus aureus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT 108..109
FT Region /label= altered_M15_cluster
FT /note= "wild-type cluster 15 sequence Glu-Tyr-
FT has been replaced by Ala-Ala to reduce
FT immunogenicity"
XX
PN EP721982-A1.
XX
PD 17-JUL-1996.
XX
PF 06-JAN-1995; 95EP-0200023.
XX
PR 06-JAN-1995; 95EP-0200023.
XX
PA (COLL/) COLLEN D J.
XX (LEUV-) LEUVEN RES & DEV VZW.
XX
PI Collen D;
XX
DR WPI; 1996-322832/33.
XX
PT New staphylokinase mutants with reduced immunogenicity - useful for
XX treating arterial thrombosis
XX
PS Example 2; Page -: 21pp; English.
XX
CC Staphylokinase derivatives showing a reduced immunogenicity as
CC compared to wild-type staphylokinase are claimed. The derivatives
CC are useful as thrombolytic agents to treat arterial thrombosis and
CC are pref. produced by eliminating at least one of the epitopes
CC indicated in the features table. The epitopes are destroyed by
CC replacing one or more amino acid residues in a charge cluster by an
CC Ala residue. Mutations are introduced using site-directed
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus
CC lysogenic strain SakSTAR. The present sequence is a specific
CC example of a staphylokinase derivative according to the invention;
CC the sequence does not appear in the specification but can be generated
CC from the wild-type staphylokinase sequence using the description given
CC in Table 1 of the disclosure.
XX
SQ Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;
Best Local Similarity 98.5%; Pred. No. 3, 9e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

DB 61 EYVEMALDATALAYKEFRVLEDPKSAKIEVYDKNKKKEETKSPFPIKGFVVPDISEHI 120
QY 121 KNPGFNLTIRKVIIEKK 136
DB 121 KNPGFNLTIRKVIIEKK 136

RESULT 45
AAM03076
ID AAM03076 standard; protein; 136 AA.
AC AAM03076;
DT 19-FEB-1997 (first entry)
XX
XX
DE Staphylokinase derivative M21 with altered charge cluster 21.
XX
XX Staphylokinase; mutant; mutelin; variant; immunogenicity; decrease;
KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.
XX
OS Staphylococcus aureus.
OS Synthetic.
FH
FT Key Location/Qualifiers
FT Region 8..10
FT /label= altered_M21_cluster
FT /note= "Wild-type cluster 21 sequence Lys-Tyr-Lys
FT has been replaced by Ala-Tyr-Ala to
FT reduce immunogenicity"

EP721982-A1.
PD 17-JUL-1996.
XX
XX
PF 06-JAN-1995; 95EP-0200023.
XX
PR 06-JAN-1995; 95EP-0200023.
XX
PA (COLL-) COLLEN D J.
PA (LEUV-) LEUVEN RES & DEV VZW.
PI Collen D;
XX
XX WPI; 1996-322832/33.
DR
XX
XX
PT New staphylokinase mutants with reduced immunogenicity - useful for
PT treating arterial thrombosis
XX
PS Example 2; Page -: 21pp; English.
XX
XX Staphylokinase derivatives showing a reduced immunogenicity as
CC compared to wild-type staphylokinase are claimed. The derivatives
CC are useful as thrombolytic agents to treat arterial thrombosis and
CC are pref. produced by eliminating at least one of the epitopes
CC indicated in the features table. The epitopes are destroyed by
CC replacing one or more amino acid residues in a charge cluster by an
CC Ala residue. Mutations are introduced using site-directed
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus
CC lysogenic strain SakSTAR. The present sequence is a specific
CC example of a staphylokinase derivative according to the invention;
CC the sequence does not appear in the specification but can be generated
CC from the wild-type staphylokinase sequence using the description given
CC in Table 1 of the disclosure.
XX
XX
SQ Sequence 136 AA.

Query Match 98.3%; Score 703; DB 17; Length 136;
Best Local Similarity 98.5%; Pred. No. 3..9e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDDGKTKKGDASVFEPTGPILMVNTGVDSKNEILSPHYVEFPPIKPGTTLTKETI 60

DB 1 SSSFDDGKTKKGDASVFEPTGPILMVNTGVDSKNEILSPHYVEFPPIKPGTTLTKETI 60
QY 61 EYVEMALDATALAYKEFRVLEDPKSAKIEVYDKNKKKEETKSPFPIKGFVVPDISEHI 120
DB 61 EYVEMALDATALAYKEFRVLEDPKSAKIEVYDKNKKKEETKSPFPIKGFVVPDISEHI 120
QY 121 KNPGFNLTIRKVIIEKK 136
DB 121 KNPGFNLTIRKVIIEKK 136

RESULT 46
AAM03095
ID AAM03095 standard; protein; 136 AA.
AC AAM03095;
DT 19-FEB-1997 (first entry)
XX
XX
DE Staphylokinase derivative M10 with altered charge cluster 10.
XX
XX Staphylokinase; mutant; mutelin; variant; immunogenicity; decrease;
KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.
XX
OS Staphylococcus aureus.
OS Synthetic.
FH
FT Key Location/Qualifiers
FT Region 86..88
FT /label= altered_M10_cluster
FT /note= "Wild-type cluster 10 sequence Lys-Ile-Glu
FT has been replaced by Ala-Ile-Ala to reduce
FT immunogenicity"

W09621016-A2.
PD 11-JUL-1996.
XX
XX
PF 03-JAN-1996; 96WO-EP00081.
XX
PR 17-NOV-1995; 95JP-0299781.
PR 06-JAN-1995; 95EP-0200023.
PR 11-JAN-1995; 95US-0371505.
PR 09-JUN-1995; 95EP-0201531.
PR 06-JUL-1995; 95US-0499032.
XX
XX
PA (COLL-) COLLEN D.
PA (LEUV-) LEUVEN RES & DEV VZW.
PI Collen D;
XX
XX WPI; 1996-333991/33.
DR
XX
XX
PT New staphylokinase derivs. having reduced immunogenicity - useful
PT for treating arterial thrombosis
XX
PS Example 2; Page -: 58pp; English.
XX
XX Staphylokinase derivatives showing a reduced immunogenicity as
CC compared to wild-type staphylokinase are claimed. The derivatives
CC are useful as thrombolytic agents to treat arterial thrombosis and
CC are pref. produced by eliminating at least one of the epitopes
CC indicated in the features table. The epitopes are destroyed by
CC replacing one or more amino acid residues in a charge cluster by an
CC Ala residue. Mutations are introduced using site-directed
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus
CC lysogenic strain SakSTAR. The present sequence is a specific
CC example of a staphylokinase derivative according to the invention;
CC the sequence does not appear in the specification but can be generated
CC from the wild-type staphylokinase sequence using the description given
CC in Table 1 of the disclosure.
XX

CC are useful as thrombolytic agents to treat arterial thrombosis and
 CC are pref. produced by eliminating at least one of the epitopes
 CC indicated in the features table. The epitopes are destroyed by
 CC replacing one or more amino acid residues in a charge cluster by an
 CC Ala residue. Mutations are introduced using site-directed
 CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus
 CC lysogenic strain SakSTAR. The present sequence is that of the
 CC specifically claimed staphylokinase derivative M3; the sequence
 CC does not appear in the specification but can be generated from the
 CC wild-type staphylokinase sequence using the description given in
 CC the claims.

XX Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;
 Best Local Similarity 98.5%; Pred. No. 3.9e-69;
 Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDKGRKKGGDASYEPTGPIAMVNTGVDSKGNELSPHYVEFPKPTLTKEKI 60
 |||||
 Db 1 sssfdkgkykkgdasyfepcpgylmwnvcgvdasgnallspkyvefpkptltkexl 60

QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVYTYDKNKKKEETKSPITTEKGFVVPDLSEHI 120
 |||||
 Db 61 eyvwealdataykefrvveldpsakievtydknkkkeetsfptekgfvpdlsehl 120

QY 121 KNPGENLITKRVYIEKK 136
 |||||
 Db 121 knpgfnliltkvylekk 136

RESULT 49
 AAW03090
 ID AAW03090 standard; protein; 136 AA.

XX AC AAW03090;
 XX DT 19-FEB-1997 (first entry)

XX DE Staphylokinase derivative M3 with altered charge cluster 3.

XX KW Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;
 XX KM derivative; SakSTAR; arterial thrombosis; thrombolytic agent.

XX OS Staphylococcus aureus.
 XX OS Synthetic.

XX FH Key Location/Qualifiers
 XX FT Region 35..38

XX FT /label= altered_M3_cluster
 XX FT /note= "wild-type cluster 3 sequence Lys-Gly-Asn-Glu
 XX FT has been replaced by Ala-Gly-Asn-Ala to
 XX FT reduce immunogenicity"

XX MO9621016-A2.

XX PN 11-JUL-1996.

XX PD 03-JAN-1996; 96MO-EP00081.

XX PR 17-NOV-1995; 95JP-0299781.

XX PR 06-JAN-1995; 95EP-0200023.

XX PR 11-JAN-1995; 95US-0371505.

XX PR 09-JUN-1995; 95EP-0201531.

XX PR 06-JUL-1995; 95US-0499092.

XX PA (COLL/) COLLEN D.
 XX PA (LEUV-) LEUVEN RES & DEV VZW.

XX PI Collen D;
 XX DR WPI; 1996-333991/33.

XX New staphylokinase derivs. having reduced immunogenicity - useful
 XX for treating arterial thrombosis
 XX PS Claim 9; Page -: 58pp; English.

XX Staphylokinase derivatives showing a reduced immunogenicity as
 XX compared to wild-type staphylokinase are claimed. The derivatives
 XX are useful as thrombolytic agents to treat arterial thrombosis and
 XX are pref. produced by eliminating at least one of the epitopes
 XX indicated in the features table. The epitopes are destroyed by
 XX replacing one or more amino acid residues in a charge cluster by an
 XX Ala residue. Mutations are introduced using site-directed
 XX mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus
 XX lysogenic strain SakSTAR. The present sequence is that of the
 XX specifically claimed staphylokinase derivative M3; the sequence
 XX does not appear in the specification but can be generated from the
 XX wild-type staphylokinase sequence using the description given in
 XX the claims.

XX Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;
 Best Local Similarity 98.5%; Pred. No. 3.9e-69;
 Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDKGRKKGGDASYEPTGPIAMVNTGVDSKGNELSPHYVEFPKPTLTKEKI 60
 |||||
 Db 1 sssfdkgkykkgdasyfepcpgylmwnvcgvdasgnallspkyvefpkptltkexl 60

QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVYTYDKNKKKEETKSPITTEKGFVVPDLSEHI 120
 |||||
 Db 61 eyvwealdataykefrvveldpsakievtydknkkkeetsfptekgfvpdlsehl 120

QY 121 KNPGENLITKRVYIEKK 136
 |||||
 Db 121 knpgfnliltkvylekk 136

RESULT 50
 AAW03088
 ID AAW03088 standard; protein; 136 AA.

XX AC AAW03088;
 XX DT 19-FEB-1997 (first entry)

XX DE Staphylokinase derivative M21 with altered charge cluster 21.

XX KW Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;
 XX KM derivative; SakSTAR; arterial thrombosis; thrombolytic agent.

XX OS Staphylococcus aureus.
 XX OS Synthetic.

XX FH Key Location/Qualifiers
 XX FT Region 8..10

XX FT /label= altered_M21_cluster
 XX FT /note= "wild-type cluster 21 sequence Lys-Tyr-Lys
 XX FT has been replaced by Ala-Tyr-Ala to
 XX FT reduce immunogenicity"

XX MO9621016-A2.

XX PN 11-JUL-1996.

XX PD 03-JAN-1996; 96MO-EP00081.

XX PR 17-NOV-1995; 95JP-0299781.

XX PR 06-JAN-1995; 95EP-0200023.

XX PR 11-JAN-1995; 95US-0371505.

XX PR 09-JUN-1995; 95EP-0201531.

PR 06-JUL-1995; 95US-0499092.

XX (COLL/) COLLEN D.

PA (LEUV-) LEUVEN RES & DEV VZW.

XX COLLEN D;

XX WPI; 1996-333991/33.

XX New staphylokinase derivs. having reduced immunogenicity - useful
PT for treating arterial thrombosis

XX Example 2; Page -: 58pp; English.

XX staphylokinase derivatives showing a reduced immunogenicity as
CC compared to wild-type staphylokinase are claimed. The derivatives
CC are useful as thrombolytic agents to treat arterial thrombosis and
CC are pref. produced by eliminating at least one of the epitopes
CC indicated in the features table. The epitopes are destroyed by
CC replacing one or more amino acid residues in a charge cluster by an
CC Ala residue. Mutations are introduced using site-directed
CC mutagenesis on wild-type staphylokinase genomic DNA from the S. aureus
CC lysogenic strain SakSTAR. The present sequence is a specific
CC example of a staphylokinase derivative according to the invention;
CC the sequence does not appear in the specification but can be generated
CC from the wild-type staphylokinase sequence using the description given
CC in Table 1 of the disclosure.

XX Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;

Best Local Similarity 98.5%; Pred. No. 3.9e-69;

Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	SSSEFDGKXKKGGDASFEPTGPIYIMVNTGVDSKGNELLSPHYVEFPPIKPGTTLKEKI	60
Db	1	sssfidkgayakgddasyfepctgpiimvntgvdskgnelispiyvefpikpgtltlkeki	60
QY	61	EYVVENALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFVVPDLSEHI	120
Db	61	eyvvenaldetaykefrvveldpsakievtydknkkkeetsfpitkfgfvyvpdlsehi	120
QY	121	KNPGENLITKVIIEKK	136
Db	121	knpqfnlftkvviek	136

Search completed: September 1, 2002, 03:17:40
Job time: 7989 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2002, 01:58:01 ; Search time 24.89 seconds
(without alignments)
133.462 Million cell updates/sec

Title: US-09-728-670-10

Perfect score: 715
1 SSSFDKGRKKKGDASYEP.....SEHKNPGFNLTQVIEKK 136

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6C-COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715	100.0	136	1 US-08-371-505-2	Sequence 2, Appl1
2	715	100.0	136	2 US-08-784-971-5	Sequence 5, Appl1
3	715	100.0	137	1 US-08-256-261-6	Sequence 6, Appl1
4	715	100.0	137	3 US-08-852-299-6	Sequence 4, Appl1
5	711	99.4	137	1 US-08-256-261-4	Sequence 4, Appl1
6	711	99.4	137	3 US-08-852-299-4	Sequence 4, Appl1
7	705	98.6	163	1 US-08-075-545-1	Sequence 1, Appl1
8	695	97.2	137	1 US-08-256-261-2	Sequence 2, Appl1
9	695	97.2	137	3 US-08-852-299-2	Sequence 17, Appl1
10	695	97.2	163	1 US-08-256-261-17	Sequence 17, Appl1
11	695	97.2	163	3 US-08-852-299-17	Sequence 14, Appl1
12	692	96.8	137	1 US-08-256-261-14	Sequence 14, Appl1
13	692	96.8	137	3 US-08-852-299-14	Sequence 14, Appl1
14	689	96.4	137	1 US-08-256-261-12	Sequence 12, Appl1
15	689	96.4	137	3 US-08-852-299-12	Sequence 12, Appl1
16	653	91.3	126	1 US-08-075-545-2	Sequence 2, Appl1
17	643	89.9	127	1 US-08-256-261-8	Sequence 8, Appl1
18	643	89.9	127	3 US-08-852-299-8	Sequence 8, Appl1
19	620	86.7	123	1 US-08-256-261-10	Sequence 10, Appl1
20	620	86.7	123	3 US-08-852-299-10	Sequence 10, Appl1
21	79	11.0	15	4 US-09-561-108-19	Sequence 19, Appl1
22	79	11.0	15	4 US-09-561-108-19	Sequence 19, Appl1
23	76	10.6	15	4 US-09-561-108-20	Sequence 20, Appl1
24	76	10.6	15	4 US-09-561-108-20	Sequence 20, Appl1
25	74	10.3	592	4 US-09-000-145-6	Sequence 6, Appl1
26	73.5	10.3	2522	4 US-09-251-645-13	Sequence 13, Appl1
27	73	10.2	14	2 US-08-934-222-23	Sequence 23, Appl1

28	73	10.2	14	2 US-08-933-402-23	Sequence 23, Appl1
29	73	10.2	14	2 US-09-207-621-23	Sequence 23, Appl1
30	73	10.2	14	2 US-08-532-818-23	Sequence 23, Appl1
31	73	10.2	14	3 US-09-231-797-23	Sequence 23, Appl1
32	73	10.2	14	3 US-08-934-224-23	Sequence 23, Appl1
33	73	10.2	14	3 US-08-933-843-23	Sequence 23, Appl1
34	73	10.2	14	4 US-08-934-223-23	Sequence 23, Appl1
35	73	10.2	14	4 US-09-413-492-23	Sequence 23, Appl1
36	73	10.2	770	1 US-08-525-654A-1	Sequence 1, Appl1
37	73	10.2	771	1 US-08-469-005A-10	Sequence 3, Appl1
38	73	10.2	2509	1 US-08-261-907-2	Sequence 10, Appl1
39	73	10.2	2511	4 US-09-339-152A-29	Sequence 29, Appl1
40	70.5	9.9	706	1 US-08-007-999B-4	Sequence 4, Appl1
41	70.5	9.9	706	2 US-08-689-276A-4	Sequence 4, Appl1
42	70.5	9.9	706	2 US-08-689-276A-4	Sequence 4, Appl1
43	69	9.7	423	1 US-08-476-008-8	Sequence 8, Appl1
44	69	9.7	423	1 US-08-306-063-8	Sequence 8, Appl1
45	69	9.7	423	1 US-08-833-485-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-371-505-2
; Sequence 2, Application US/08371505
; Patent No. 5635754
; GENERAL INFORMATION:
; APPLICANT: COLLEN, DESIRE
; TITLE OF INVENTION: STAPHYLOKINASE DERIVATIVES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: WEBB ZIESENHEIM BREUNING LOGSDON ORKIN & HANSON, P.C.
; STREET: 700 KOPPELS BUILDING, 436 SEVENTH AVENUE
; CITY: PITTSBURGH
; STATE: PENNSYLVANIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 15222-2363
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; OPERATING SYSTEM: DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/371,505
; FILING DATE: 11 JAN 1995
; CLASSIFICATION: 424
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 136
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: UNKNOWN
; US-08-371-505-2

Query Match 100.0%; Score 715; DB 1; Length 136;
Best Local Similarity 100.0%; Pred No. 2e-78;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSFDKGRKKKGDASYEPPTGYLWVNVGVDSKGNELLSPHYVEFP1KPGTTLTKRK1 60
DB 1 SSSFDKGRKKKGDASYEPPTGYLWVNVGVDSKGNELLSPHYVEFP1KPGTTLTKRK1 60
QY 61 EYVEALMDTAKKRRVYVLDPSAKIEVYIYDKNNKKEETKSPITTEKGFVVPDISEH1 120
DB 61 EYVEALMDTAKKRRVYVLDPSAKIEVYIYDKNNKKEETKSPITTEKGFVVPDISEH1 120
QY 121 KNGFNLITKVIEKK 136
DB 121 KNGFNLITKVIEKK 136

RESULT 2
US-08-784-971-5
Sequence 5, Application US/08784971
Patent No. 5951980
GENERAL INFORMATION:
APPLICANT: COLLEN, DESIRE
TITLE OF INVENTION: NEW STAPHYLOKININASE DERIVATIVES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: THE WEBB LAW FIRM
STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE
CITY: PITTSBURGH
STATE: PENNSYLVANIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 15219-1818
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISK
COMPUTER: DIGITAL VENTURIS GL 6200
OPERATING SYSTEM: DOS
SOFTWARE: MICROSOFT WORD 2.0c
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,971
FILING DATE: 16-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/371,505
FILING DATE: 11-JAN-1995
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 136
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: UNKNOWN
US-08-784-971-5

Query Match 100.0%; Score 715; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 2e-78;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSSFDKGYKKGGDASFEPTGPLYLVNVTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 60
Db 1 SSSFDKGYKKGGDASFEPTGPLYLVNVTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 60
Qy 61 EYVEWALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
Db 61 EYVEWALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
Qy 121 KNPGFNLITKVIYEKK 136
Db 121 KNPGFNLITKVIYEKK 136

RESULT 3
US-08-256-261-6
Sequence 6, Application US/08256261
Patent No. 5801037
GENERAL INFORMATION:
APPLICANT: Behnke, Detlef
APPLICANT: Schlotz, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Helinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-256-261-6

Query Match 100.0%; Score 715; DB 1; Length 137;
Best Local Similarity 100.0%; Pred. No. 2e-78;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSSFDKGYKKGGDASFEPTGPLYLVNVTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 60
Db 2 SSSFDKGYKKGGDASFEPTGPLYLVNVTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 61
Qy 61 EYVEWALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
Db 62 EYVEWALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 121
Qy 121 KNPGFNLITKVIYEKK 136
Db 122 KNPGFNLITKVIYEKK 137

RESULT 4
US-08-852-299-6
Sequence 6, Application US/08852299
Patent No. 6010897
GENERAL INFORMATION:
APPLICANT: Behnke, Detlef
APPLICANT: Schlotz, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Helinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,299
FILING DATE: 17-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/256,261
FILING DATE:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-852-299-6

Query Match	100.0%;	Score 715;	DB 3;	Length 137;
Best Local Similarity	100.0%;	Pred. No. 2e-78;		
Matches 136; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

OY 1 SSSDCKGKYYKGGDDASYEPTGTGLMNVNTGVDSKGNELLSPHYVEFPKPGITLTKEKI 60
Db 2 SSSEDKGKYKKGGDDASYEPTGTGLMNVNTGVDSKGNELLSPHYVEFPKIPGTTLTKEKI 61

QY 61 EYVEWALDATALYKKEERVVELDPSAKIEVTYYDKNKKKEETKSPITTEKGVVPPDLSEHI 120
 |||||
 Db 62 EYVEWALDATALYKKEERVVELDPSAKIEVTYYDKNKKKEETKSPITTEKGVVPPDLSEHI 121

Oy	121	KNPGFNLTQVIEKK	136
Db	122	KNPGFNLTQVIEKK	137

RESULT 5
US-08-256-261-4
; Sequence 4, Application US/08256261
Date: 08/03/07

```

: GENERAL INFORMATION:
: APPLICANT: Behnke, Detlef
: APPLICANT: Schlotz, Bernhard
: APPLICANT: Albrecht, Sybille
: APPLICANT: G hrs, Karl-Heinz
: APPLICANT: Hartmann, Manfred
: TITLE OF INVENTION: Expression of signal-peptide-free
: TITLE OF INVENTION: staphylokinases
: NUMBER OF SEQUENCES: 40

```

ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:

Query Match	99.4%	Score 711;	DB 1;	length 137;
Best Local Similarity	99.3%	Pred. No. 6.1e-78;		
Matches 135; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

Qy 1 SSSFPKGGKKKGGDDASFEPTGPLMVNNTGVDSKGNELSPHYVEFPKPGTTLTKKI 600
|||||
2 SSSFPKGGKKKGGDDASFEPTGPLMVNNTGVDSKGNELSPHYVEFPKPGTTLTKKI 611

[illegible]

QY	121	KNPGFNLITKVIEKK	136
Db	122	KNPGFNLITKVIEKK	137

RESULT 6
US-08-852-299-4
; Sequence 4, Application US/08852299
; Patent No. 6010897

APPLICANT: Behnke, Detlef
APPLICANT: Schlotz, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal peptide-free
TITLE OF INVENTION: staphylokinases
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:

COMPUTER READABLE FORM:

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,299
FILING DATE: 17-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/256,261

Query Match	99.48;	Score 711;	DB 3;	Length 137;
Best Local Similarity	99.38;	Pred. No. 6.1e-78;		
Matches 135; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy 1 SSSFDKGGKKRGDDASIEPTGPIYLMVNTGVDSKGNELLSPHYVEFPRIKPGTITLKEKI 60
Db 2 SSSEFDKGGKKRGDDASYEPTGPILMVNTGVDSKGNELLSPHYVEFPRIKPGTITLKEKI 61

61 EYVEMALDQAYKERRVELDPSAKIEVTYYDKNKKKEETKSPFITEKGFVPPDLSEHI 120
 |||||
 62 EYVEMALDQAYKERRVELDPSAKIEVTYYDKNKKKEETKSPFITEKGFVPPDLSEHI 121

QY	121	KNPGFNLITKVIIEKK	136
Db	122	KNPGFNLITKVIIEKK	137

RESULT 7
 US-08-075-545-1
 Sequence 1, Application US/08075545
 Patent No. 5475089
 GENERAL INFORMATION:
 APPLICANT: MATSUDO, Osamu; SAKAI, Masashi; SHIMIDU,
 Katsuku; SANSABA, Hiroshi; WATANABE,
 APPLICANT: Tsunekazu; MATSUMOTO, Tsuneo; SHISHIDO,
 APPLICANT: Yoshiyuki; HASHIMOTO, Shunsuke; YOKOKURA
 APPLICANT: Tetsuo; ONEUE, Masaharu; SAKO, Tomoyuki
 TITLE OF INVENTION: THROMBOLYTIC AGENT
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:

```

? ADDRESS: Fish & Neave
? STREET: 1251 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: USA
? ZIP: 10020
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
?
```

```

CORRESPONDENCE: ADDRESS:
ADDRESSSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,299
FILING DATE: 17-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/256,261

FILING DATE:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-852-299-2

```


STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-256-261-14

Query Match 96.8%; Score 692; DB 1; Length 137;
Best Local Similarity 97.1%; Pred. No. 1.2e-75;
Matches 132; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 SSSFDGKRRKKDDASYEPTGPLYMNVNTGVDKGNELLSPHYVEPIKPGTTLTKRKI 60
DB 2 SSSFDGKRRKKDDASYEPTGPLYMNVNTGVDKGNELLSPHYVEPIKPGTTLTKRKI 61
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFFVVDLSEHI 120
DB 62 EYVEMALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFFVVDLSEHI 121
QY 121 KNPGFNLTIKVYIEKK 136
DB 122 KNPGFNLTIKVYIEKK 137

RESULT 13
US-08-852-299-14
Sequence 14, Application US/08852299
Patent No. 6010897
GENERAL INFORMATION:
APPLICANT: Behnke, Detlef
APPLICANT: Schlotz, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,299
FILING DATE: 17-MAY-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/256,261
FILING DATE:
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-852-299-14

Query Match 96.8%; Score 692; DB 3; Length 137;
Best Local Similarity 97.1%; Pred. No. 1.2e-75;
Matches 132; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 SSSFDGKRRKKDDASYEPTGPLYMNVNTGVDKGNELLSPHYVEPIKPGTTLTKRKI 60
DB 2 SSSFDGKRRKKDDASYEPTGPLYMNVNTGVDKGNELLSPHYVEPIKPGTTLTKRKI 61
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFFVVDLSEHI 120
DB 62 EYVEMALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFFVVDLSEHI 121
QY 121 KNPGFNLTIKVYIEKK 136
DB 122 KNPGFNLTIKVYIEKK 137

RESULT 14
US-08-256-261-12
Sequence 12, Application US/08256261
Patent No. 5801037
GENERAL INFORMATION:
APPLICANT: Behnke, Detlef
APPLICANT: Schlotz, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-256-261-12

Query Match 96.4%; Score 689; DB 1; Length 137;
Best Local Similarity 97.1%; Pred. No. 2.7e-75;
Matches 132; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 SSSFDGKRRKKDDASYEPTGPLYMNVNTGVDKGNELLSPHYVEPIKPGTTLTKRKI 60
DB 2 SSSFDGKRRKKDDASYEPTGPLYMNVNTGVDKGNELLSPHYVEPIKPGTTLTKRKI 61
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFFVVDLSEHI 120

Db 62 EYVEMALDANAYEERVELDPASAKIEVYYDANKKKKEETKSPPIETKGVVPDLSEHI 121
QY 121 KNPGFNLTIKVIEKK 136
122 KNPENLTITKVIIEKK 137
Db 122 KNPGFNLTITKVIIEKK 137

RESULT 15
US-08-852-299-12
; Sequence 12, Application US/08852299
; Patent No. 6010897

Query Match	96.48	Score 689	DB 3	Length 137
Best Local Similarity	97.18	Pred. NO. 2.7e-75		
Matches 132; Conservative	0	Mismatches 4	Indels 0	Gaps 0

```
Qy      61 EYYEMALDATAYKEFRVVELDPSAKIEYTYIDKNNKKKEETKSPTTEKGTVVPDISEHI   1200
          |||||
Db      62 EYYEMALDATAYAKEFRVVELDPSAKIEVTYIDKNKKKEETKSFTTEKGFVPPDLSEHI   1211
```

Qy	121	KNPGFNLT	KVIEKK	136
Db	122	KNPGFNLT	KVIEKK	137

RESULT 16
US-08-075-545-2
; Sequence 2, Application US/08075545
; Patent No. 5475089
GENERAL INFORMATION:
; APPLICANT: MATSUO, Osamu; SAKAI, Masashi; SHIMURA,
; APPLICANT: KISAKU; SANJAMA, Hiroshi; WATANABE,
; APPLICANT: TSUNEKAZU; MATSUMOTO, Tsuneo; SHISHIDO,

```

1  APPLICANT:  Yoshituki, Hashimoto, Shusuke, Yokokura,
2  APPLICANT:  Teruo, Onoue, Maeharu, Sako, Tomoyuki
3  TITLE OF INVENTION:  THROMBOLYTIC AGENT
4  NUMBER OF SEQUENCES:  2
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE:  KECK, MAHIN & CATE
7  STREET:  P. O. BOX 06110
8  CITY:  CHICAGO
9  STATE:  ILLINOIS
10 COUNTRY:  U.S.A.
11 ZIP:  60606-0110
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE:  Floppy disk
14 COMPUTER:  IBM PC compatible
15 OPERATING SYSTEM:  MS-DOS/PC DOS
16 SOFTWARE:  DOS Text
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER:  US/08/075,545
19 FILING DATE:  14-JUN-1993
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER:  PCT/JP91/01722
22 FILING DATE:  17-DEC-1991
23 ATTORNEY/AGENT INFORMATION:
24 NAME:  Fleit, Martin, Gollin, Michael A.
25 REGISTRATION NUMBER:  16,900; 31,957
26 REFERENCE/DOCKET NUMBER:  47004-015
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE:  (202) 789-3400
29 TELEFAX:  (202) 789-1158
30 INFORMATION FOR SEQ ID NO:  2:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH:  126 amino acids
33 TYPE:  amino acid
34 STRANDEDNESS:  unknown
35 TOPOLOGY:  linear
36 MOLECULE TYPE:  peptide
37 US-08-075-545-2

```

Query Match	91.3%	Score 653	DB 1	Length 126
Best Local Similarity	98.4%	Pred. No. 5.1e-71		
Matches 124	Conservative	0	Mismatches 2	Indels 0
				Gaps 0

Qy 71 TAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEGEFVPDLSEHKPNPGFNILTK 130
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEGEFVPDLSEHKPNPGFNILTK 120

QY	131	VVIEKK	136
Db	121	VVIEKK	126

RESULT 17
US-08-256-261-8
: Sequence 8, Application US/08256261
: Patent No. 5801037
: GENERAL INFORMATION:
: APPLICANT: Behnke, Detlef
: APPLICANT: Schloft, Bernhard
: APPLICANT: Albrecht, Sybille
: APPLICANT: G hrs, Karl-Heinz
: APPLICANT: Hartmann, Manfred
: TITLE OF INVENTION: Expression of signal-peptide-free
: TITLE OF INVENTION: staphylokinases
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Neave
: STREET: 1251 Avenue of the Americas
: CITY: New York

STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-256-261-8

Query Match 89.9%; Score 643; DB 1; Length 127;
Best Local Similarity 97.6%; Pred. No. 8.3e-70;
Matches 123; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 KGDDASYEPTGPLYLVNTGVDSKGNELLSPHYVEFPKPGTTLTKKIEYVEMALDA 70
|||||
DB 2 KGDDASYEPTGPLYLVNTGVDSKGNELLSPHYVEFPKPGTTLTKKIEYVEMALDA 61

QY 71 TAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSHINKPGFNLTIK 130
|||||
DB 62 TAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSHINKPGFNLTIK 121

QY 131 VVIEKK 136
|||||
DB 122 VVIEKK 127

RESULT 18
US-08-852-299-8
Sequence 8, Application US/08852299
Patent No. 6010897
GENERAL INFORMATION:
APPLICANT: Behnke, Detlef
APPLICANT: Schlotz, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,299
FILING DATE: 17-MAY-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/256,261
FILING DATE:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-852-299-8

Query Match 89.9%; Score 643; DB 3; Length 127;
Best Local Similarity 97.6%; Pred. No. 8.3e-70;
Matches 123; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 KGDDASYEPTGPLYLVNTGVDSKGNELLSPHYVEFPKPGTTLTKKIEYVEMALDA 70
|||||
DB 2 KGDDASYEPTGPLYLVNTGVDSKGNELLSPHYVEFPKPGTTLTKKIEYVEMALDA 61

QY 71 TAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSHINKPGFNLTIK 130
|||||
DB 62 TAYKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSHINKPGFNLTIK 121

QY 131 VVIEKK 136
|||||
DB 122 VVIEKK 127

RESULT 19
US-08-256-261-10
Sequence 10, Application US/08256261
Patent No. 5801037
GENERAL INFORMATION:

APPLICANT: Behnke, Detlef
APPLICANT: Schlotz, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:

CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-256-261-10

Query Match 86.7%; Score 620; DB 1; Length 123;
Best Local Similarity 97.5%; Pred. No. 4.6e-67;
Matches 119; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 ASYFEPTGPLYLVNTGVDSKGNELLSPHYVEFPKPGTTLTKKIEYVEMALDAATAYK 74
|||||

DB 2 ASYFEPTGPLYLVNTGVDSKGNELLSPHYVEFPKPGTTLTKKIEYVEMALDAATAYK 61

QY 75 EFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSHINKPGFNLTIKVIE 134
|||||
DB 62 EFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSHINKPGFNLTIKVIE 121

OY 135 KK 136
11
DB 122 KK 123

RESULT 20

US-08-852-299-10
; Sequence 10, Application US/08852299
; Patent No. 6010897
; GENERAL INFORMATION:
; APPLICANT: Behke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; TITLE OF INVENTION: staphylokinases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,299
; FILING DATE: 17-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,261
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 123 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-852-299-10

Query Match 86.7%; Score 620; DB 3; Length 123;

Best Local Similarity 97.5%; Pred. No. 4.6e-67;
Matches 119; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 15 ASYFPTGTYLWNTGYVDSKGNELSPHYVEPIKPGTTLTKKEIEYVEMALDATAK 74
|||||
DB 2 ASYFPTGTYLWNTGYVDSKGNELSPHYVEPIKPGTTLTKKEIEYVEMALDATAK 61
|||||
OY 75 EFRVVELDPSAKIEVYTDKKNKKKEETSPITEKGFVVPDLSEHIKNPGENLTIVIE 134
|||||
DB 62 EFRVVELDPSAKIEVYTDKKNKKKEETSPITEKGFVVPDLSEHIKNPGENLTIVIE 121
|||||

OY 135 KK 136
11
DB 122 KK 123

RESULT 21

US-09-561-500-19
; Sequence 19, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brecken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500

; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; US-09-561-500-19

Query Match 11.0%; Score 79; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSSFDKGRKKKGDDA 15
|||||
DB 1 SSSFDKGRKKKGDDA 15

RESULT 22

US-09-561-108-19
; Sequence 19, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brecken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VE
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; US-09-561-108-19

Query Match 11.0%; Score 79; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSSFDKGRKKKGDDA 15
|||||
DB 1 SSSFDKGRKKKGDDA 15

RESULT 23

US-09-561-500-20
; Sequence 20, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brecken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentn Ver. 2.0

```
; SEQ ID NO 20
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-20
```

```
Query Match          10.6%; Score 76; DB 4; Length 15;
Best Local Similarity 93.3%; Pred. No. 0.0057;
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 SSSFDRGKRYKRGDDA 15
| | | | | | | | | | | | | | |
Db 1 SSSFDRGKRYKRGDDA 15

RESULT 24
US-09-561-108-20
; Sequence 20, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-20
```

```
Query Match          10.6%; Score 76; DB 4; Length 15;
Best Local Similarity 93.3%; Pred. No. 0.0057;
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 SSSFDRGKRYKRGDDA 15
| | | | | | | | | | | | | | |
Db 1 SSSFDRGKRYKRGDDA 15
```

```
RESULT 25
US-09-000-145-6
; Sequence 6, Application US/09000145
; Patent No. 6169172
; GENERAL INFORMATION:
; APPLICANT: DEVAUCHELLE, Gerard
; APPLICANT: GARNIER, Laurence
; APPLICANT: CAHOREAU, Claire
; APPLICANT: CERUTTI, Martine
; TITLE OF INVENTION: USE OF A PROLACTIN RECEPTOR OR GROWTH HORMONE RECEPTOR
; TITLE OF INVENTION: INTRACYTOLASMIC DOMAIN FOR ACHIEVING PROTEIN SECRETION
; FILE REFERENCE: 0384-0047-0XPCT
; CURRENT APPLICATION NUMBER: US/09/000,145
; CURRENT FILING DATE: 1998-03-16
; EARLIER APPLICATION NUMBER: PCT/FR96/01237
; EARLIER FILING DATE: 1996-08-02
; EARLIER APPLICATION NUMBER: FR 95/09420
; EARLIER FILING DATE: 1995-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 6
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-000-145-6
```

```
Query Match          10.3%; Score 74; DB 4; Length 592;
Best Local Similarity 23.9%; Pred. No. 2.3;
Matches 34; Conservative 17; Mismatches 43; Indels 48; Gaps 7;
```

```
OY 4 FDKCKRYKGGDAST-----PEPTGPR--LMVNTGVDSKGNELSP-HYVEFP--I 49
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 256 FDNHLEKGRSEELLALGCODEPPTSDCEBLLVFEFLVDNEDERLMPSHSKRYPCGV 315

OY 50 KPGTTLTKREIEYVWALDATAKKEFRVVELDPSAKIEVYYDK-----NKKKEETKSP 105
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 316 KP-----THLDPDSGSGHSYSHLSLSEKCEPQAYP 348

OY 106 ITEKGFVPDLSEHIKNPGFNL 127
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 349 PT---LHPEITEKPEMPEANI 367
```

```
RESULT 26
US-09-251-645-13
; Sequence 13, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
; FILE REFERENCE: GCG1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-13
```

```
Query Match          10.3%; Score 73.5; DB 4; Length 2522;
Best Local Similarity 28.9%; Pred. No. 23;
Matches 37; Conservative 16; Mismatches 42; Indels 33; Gaps 8;
```

```
OY 12 GDD---ASYPEPTGPRVLMVNTGVDSK-----NELSPHYVEFPKRGCT---T 54
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1461 GDDYATDSLKPNDLKRYVT--DSKGTATDVSGPYDINTAISPAVQYTVKRGSEQT 1518

OY 55 LTKER-IEYVWALDATAKKEFRVVELD-----PSAKIEVY-----YDKNKKKEET 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1519 FTADKNVSIQSPSPFDEMN--QFNALIEDSSLNFTNNSASIDITFAFAEDKRLGYES 1577

OY 102 KSPITEK 109
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1578 FSPITRK 1585
```

```
RESULT 27
US-08-934-222-23
; Sequence 23, Application US/08934222
; Patent No. 5928896
; GENERAL INFORMATION:
; APPLICANT: EVANS, Herbert J.
```

APPLICANT: KINI, R. Manjunatha
TITLE OF INVENTION: Polypeptides That Include Conformation-
CONSTRAINING GROUPS WHICH FLANK A PROTEIN-PROTEIN INTERACTION
TITLE OF INVENTION: Site
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: Suite 500, 3000 K Street NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20007
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,222
FILING DATE: 19-SEPT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/532,818
FILING DATE: 03-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/143,364
FILING DATE: 29-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/051,741
FILING DATE: 23-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Isaacson, John P.
REGISTRATION NUMBER: 33,751
REFERENCE/DOCKET NUMBER: 040433/0148
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-934-222-23

Query Match 10.2%; Score 73; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.012;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 KGDASYEFPPTGY 24
DB 1 KPDDASYEFPPTGY 14

RESULT 28
US-08-933-402-23
Sequence 23, Application US/08933402
Patent No. 5948887
GENERAL INFORMATION:
APPLICANT: EVANS, Herbert J.
TITLE OF INVENTION: Polypeptides That Include Conformation-
CONSTRAINING GROUPS WHICH FLANK A PROTEIN-PROTEIN INTERACTION
TITLE OF INVENTION: Site
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: Suite 500, 3000 K Street NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20007
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,402
FILING DATE: 19-SEPT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/532,818
FILING DATE: 03-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/143,364
FILING DATE: 29-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/051,741
FILING DATE: 23-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Isaacson, John P.
REGISTRATION NUMBER: 33,751
REFERENCE/DOCKET NUMBER: 040433/0148
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-933-402-23

Query Match 10.2%; Score 73; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.012;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 KGDASYEFPPTGY 24
DB 1 KPDDASYEFPPTGY 14

RESULT 29
US-09-207-621-23
Sequence 23, Application US/09207621
Patent No. 5952465
GENERAL INFORMATION:
APPLICANT: EVANS, Herbert J.
TITLE OF INVENTION: Polypeptides That Include Conformation-
CONSTRAINING GROUPS WHICH FLANK A PROTEIN-PROTEIN INTERACTION
TITLE OF INVENTION: Site
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: Suite 500, 3000 K Street NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20007
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/207,621
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/532,818
FILING DATE: 03-MAY-1996
APPLICATION NUMBER: PCT/US94/04294
FILING DATE: 21-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/143,364
FILING DATE: 29-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/051,741
FILING DATE: 23-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Isaacson, John P.
REGISTRATION NUMBER: 33,751
REFERENCE/DOCKET NUMBER: 040433/0148

;; INFORMATION FOR SEQ ID NO: 23:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 14 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
US-09-207-621-23

Query Match 10.2%; Score 73; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.012;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 KPDASYFEPTGPY 24
| | | | | | | | | | | | | | | |
DB 1 KPDASYFEPTGPY 14

RESULT 30
US-08-532-818-23

;; Sequence 23, Application US/08532818
;; Patent No. 5965698
;; GENERAL INFORMATION:

;; APPLICANT: EVANS, Herbert J.
;; TITLE OF INVENTION: Polypeptides That Include Conformation-
;; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction
;; NUMBER OF SEQUENCES: 153
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: Suite 500, 3000 K Street NW
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20007

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/532,818
;; FILING DATE: 03-MAY-1996

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/04294

;; FILING DATE: 21-APR-1994
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: U.S. 08/143,364
;; FILING DATE: 29-OCT-1993

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: U.S. 08/051,741

;; FILING DATE: 23-APR-1993
;; ATTORNEY/AGENT INFORMATION:

;; NAME: Isacson, John P.
;; REGISTRATION NUMBER: 33,751

;; REFERENCE/DOCKET NUMBER: 040433/0148
;; INFORMATION FOR SEQ ID NO: 23:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 14 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear

US-08-532-818-23

Query Match 10.2%; Score 73; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.012;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 KPDASYFEPTGPY 24
| | | | | | | | | | | | | | | |
DB 1 KPDASYFEPTGPY 14

RESULT 31
US-09-231-797-23

;; Sequence 23, Application US/09231797
;; Patent No. 6084066
;; GENERAL INFORMATION:

;; APPLICANT: EVANS, Herbert J.
;; TITLE OF INVENTION: Polypeptides That Include Conformation-
;; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interactio
;; NUMBER OF SEQUENCES: 153
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: Suite 500, 3000 K Street NW
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20007

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/231,797
;; FILING DATE:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/532,818

;; FILING DATE: 03-MAY-1996
;; APPLICATION NUMBER: PCT/US94/04294

;; FILING DATE: 21-APR-1994
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: U.S. 08/143,364
;; FILING DATE: 29-OCT-1993

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: U.S. 08/051,741

;; FILING DATE: 23-APR-1993
;; ATTORNEY/AGENT INFORMATION:

;; NAME: Isacson, John P.
;; REGISTRATION NUMBER: 33,751

;; REFERENCE/DOCKET NUMBER: 040433/0148
;; INFORMATION FOR SEQ ID NO: 23:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 14 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear

US-09-231-797-23

Query Match 10.2%; Score 73; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.012;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 KPDASYFEPTGPY 24
| | | | | | | | | | | | | | | |
DB 1 KPDASYFEPTGPY 14

RESULT 32
US-08-934-224-23

;; Sequence 23, Application US/08934224
;; Patent No. 6100044
;; GENERAL INFORMATION:

;; APPLICANT: EVANS, Herbert J.
;; TITLE OF INVENTION: Polypeptides That Include Conformation-
;; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interacti
;; NUMBER OF SEQUENCES: 153
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: Suite 500, 3000 K Street NW
;; CITY: Washington
;; STATE: DC

US-09-261-907-2

Query Match	10.2%;	Score 73;	DB 4;	Length 2511;
Best Local Similarity	34.4%;	Pred. NO. 26;		
Matches	21;	Conservative	10;	Mismatches 28;
				Indels 2;
				Gaps 1.

QY	27	VNTGVGSKGNELLSPHYVEFPKPGTTLTEKEIEYVENALDNTAKERVVELDPSAK	86
	803	LHLSSIDANPNALPP--VEFPADRGFPILSPLIKMDHSLAMDVPAADPNPGSGSPISAA	860
QY	87	I 87	
	861	I 861	

```

1 RESULT 40
2 US-08-339-152A-29
3 Sequence 29 Application US/08339152A
4 Patent No. 5643726
5 GENERAL INFORMATION:
6 APPLICANT: Tanzi, Rudolph E.
7 APPLICANT: Kovacs, Dora M.
8 TITLE OF INVENTION: Methods For Modulating Transcription
9 TITLE OF INVENTION: From The Amyloid -Protein Precursor (App) Promote
10 NUMBER OF SEQUENCES: 33
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C.
13 STREET: 1100 New York Ave., NW, Suite 600
14 CITY: Washington
15 STATE: DC
16 COUNTRY: USA
17 ZIP: 20005
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patentin Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/339,152A
25 FILING DATE: 10-NOV-1994
26 CLASSIFICATION: 435
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Steffe, Eric K.
29 REGISTRATION NUMBER: 36,688
30 REFERENCE/DOCKET NUMBER: 0609,4120000
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: 202-371-2600
33 TELEFAX: 202-371-2540
34 TELEX:
35 INFORMATION FOR SEQ ID NO: 29:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 706 amino acids
38 TYPE: amino acid
39 TOPOLOGY: linear
40 US-08-339-152A-29

```

```
Query Match          9.9% Score 70.5; DB 1; Length 706;
Best Local Similarity 23.3%; Pred. No. 8;
Matches 27; Conservative 18; Mismatches 56; Indels 15; Gaps 3

QY      11 KGDASYFEPTGPGYLNVNVTGVDSKGNELLSPHYVEFPFKPQTLLTKIEIYYEMALDA 70
       ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      280 KGDDYNENENTLP-----GSDGTMSDKIELHVDKVPPLP--LPIINDVYFETISADD 329
                                     *~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*

QY      71 TAYEERNLVEDPSAKIEVTYYDKNKKKE-----TKSEPTEKGFFVVDLSHIK 121
       |:| : |:| : | : | : | : | : | : | : | : | : | : | : | : |
Db      330 NEHAFQKAKEQLIERHRNRMDRYKKKEKEAEIAQKNLPAKEROTLIHQFOAMVK 385
```

```

: Sequence 4 Application US/08007999B
: Patent No. 5851787
:
: GENERAL INFORMATION:
:   APPLICANT: Masco, Wilma
:   APPLICANT: Bupp, Keith
:   APPLICANT: Magendantz, Margaret
:   APPLICANT: Tanzl, Rudolph
:   APPLICANT: Solomon, Frank
:   TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
:   TITLE OF INVENTION: THEREOF
:   NUMBER OF SEQUENCES: 19
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
:     STREET: 1100 New York Ave., NW
:     CITY: Washington
:     STATE: DC
:     COUNTRY: USA
:     ZIP: 20005
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentln Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/007,999B
:   FILING DATE: 21-JAN-1993
:
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US 07/872,642
:   FILING DATE: 20-APR-1992
:
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US 07/930,022
:   FILING DATE: 17-AUG-1992
:
: ATTORNEY/AGENT INFORMATION:
:   NAME: Townsend, G. Kevin
:   REGISTRATION NUMBER: 34,033
:   REFERENCE/DOCKET NUMBER: 0609,3520002/JAG/GKT
:   TELECOMMUNICATION INFORMATION:
:     TELEPHONE: (202)371-2571
:     TELEFAX: (202)371-2540
:
: INFORMATION FOR SEQ ID NO: 4:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 706 amino acids
:     TYPE: amino acid
:     TOPOLOGY: linear
:
: US-08-007-999B-4
:
: Query Match          9.9%; Score 70.5; DB 2; Length 706;
: Best Local Similarity 23.3%; Pred. No. 8;
: Matches 27; Conservative 18; Mismatches 56; Indels 15; Gaps 3;
:
: Oy      11 KGDASIEYPPGPIYAMVNTGVDSGNELSPHYEPIKGGTTLTKRKIEYEWALDA 70
:           ||| : || : : : : : : : : : : : : : : : : : : : : : :
: Db      280 KGDVNEPNEP-----GSDGTMSDKETHDVKVPPT---LPTNDVYFETSAD 329
:
: Oy      71 TAYKFRVVEVDPASAKIEVTYYDKKKKEE-----TKSFPIREKGFVVPDSEHK 121
:           : | : | : | : | : | : | : | : | : | : | : | : | :
: Db      330 NEHAFQAKERQGLIERHRNRMDRVKKEWEAEALQAKNLPKAEROTLLQHOAMK 385
:
: RESULT 42
: US-08-689-276A-4
:   Sequence 4, Application US/08689276A
:   Patent No. 5891991
:
: GENERAL INFORMATION:
:   APPLICANT: Masco, Wilma
:   APPLICANT: Bupp, Keith
:   APPLICANT: Magendantz, Margaret
:   APPLICANT: Tanzl, Rudolph
:   APPLICANT: Solomon, Frank
:   TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
:   TITLE OF INVENTION: THEREOF
:   NUMBER OF SEQUENCES: 19

```


APPLICATION NUMBER: US/08/306,063
FILING DATE: 13-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/749,611
FILING DATE: 28-AUG-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/576,537
FILING DATE: 31-AUG-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hoerner Jr., Dennis R.
REGISTRATION NUMBER: 30,914
REFERENCE/DOCKET NUMBER: 38-21(10660)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)537-6099
TELEFAX: (314)537-6047
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-306-063-8

Query Match 9.7%; Score 69; DB 1; Length 423;
Best Local Similarity 24.5%; Pred. No. 5.7;
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

QY 11 KGDDASYEPTGPLYMNVTVGVDSKGNELSPHYVEPIK-----PGTTLTKERIEY 63
DB 190 KGDLYS-----KPYIDTLNLMKTFGEIENQHQQFVVGCGOSYOSPT-----YL 236
QY 64 VEMALDATAYKEFRVVELDPSAKIEVTVYDNKKKEETKSPITEK 109
DB 237 VEG--DASSASYFLAAAIKGTGTVKVGIGRNSMGDIRFADVLEK 280

RESULT 45
US-08-833-485-8
Sequence 8, Application US/08833485
Patent No. 5804425
GENERAL INFORMATION:
APPLICANT: Barry, Gerard F.
APPLICANT: Kishore, Ganesh M.
APPLICANT: Padgett, Stephen R.
APPLICANT: Stallings, William C.
TITLE OF INVENTION: Glycosate Tolerant
TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF
STREET: 700 Chesterfield Village Parkway
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833,485
FILING DATE: 07-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,063
FILING DATE: 13-SEP-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/749,611
FILING DATE: 28-AUG-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/576,537
FILING DATE: 31-AUG-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hoerner Jr., Dennis R.
REGISTRATION NUMBER: 30,914
REFERENCE/DOCKET NUMBER: 38-21(15117)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)737-6099
TELEFAX: (314)737-6047
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-833-485-8

Query Match 9.7%; Score 69; DB 1; Length 423;
Best Local Similarity 24.5%; Pred. No. 5.7;
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

QY 11 KGDDASYEPTGPLYMNVTVGVDSKGNELSPHYVEPIK-----PGTTLTKERIEY 63
DB 190 KGDLYS-----KPYIDTLNLMKTFGEIENQHQQFVVGCGOSYOSPT-----YL 236
QY 64 VEMALDATAYKEFRVVELDPSAKIEVTVYDNKKKEETKSPITEK 109
DB 237 VEG--DASSASYFLAAAIKGTGTVKVGIGRNSMGDIRFADVLEK 280

RESULT 46
US-09-137-440-8
Sequence 8, Application US/09137440
Patent No. 6248876
GENERAL INFORMATION:
APPLICANT: Barry, Gerard F.
APPLICANT: Kishore, Ganesh M.
APPLICANT: Padgett, Stephen R.
APPLICANT: Stallings, William C.
TITLE OF INVENTION: Glycosate Tolerant
TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF
STREET: 700 Chesterfield Village Parkway
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/137,440
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/833,485
FILING DATE: 07-APR-1997
APPLICATION NUMBER: US 08/306,063
FILING DATE: 13-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/749,611

;; FILING DATE: 28-AUG-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/576,537
;; FILING DATE: 31-AUG-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hoerner Jr., Dennis R.
;; REGISTRATION NUMBER: 30,914
;; REFERENCE/DOCKET NUMBER: 38-21(15117)A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (314)737-6099
;; TELEFAX: (314)737-6047
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 423 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-137-440-8

Query Match 9.7%; Score 69; DB 4; Length 423;
Best Local Similarity 24.5%; Pred. No. 5.7;
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

OY 11 KGDDASYEPPTGYLMVNTGVDSKGNELSPHYVEPIK-----PGTTLTKKEIEY 63
DB 190 KGDLVS-----KPYIDITLNLMTKTFGEIENQHQQFVVGKGSYQSPGT-----YL 236
OY 64 VEMALDATAKKEFRVVELDPSAKIEVYYDDKKRKEETKSPITEK 109
DB 237 VEG--DASSASYFLAAAIKGGTVKVTGIGRNSMQGDIRFADVLEK 280

RESULT 47

PCT-US91-06148A-8
Sequence 8, Application PC/TUS9106148A
GENERAL INFORMATION:
APPLICANT: Barry, Gerard F.
APPLICANT: Kishore, Ganesh M.
APPLICANT: Padgett, Stephen R.
TITLE OF INVENTION: Glycosate Tolerant
TITLE OF INVENTION: 5-enolpyruvylshikimate-3-phosphate Synthases
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. B94F
STREET: 700 Chesterfield Village Parkway
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/06148A
FILING DATE: 19910828
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/749,611
FILING DATE: 28-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/576,537
FILING DATE: 31-AUG-1990
ATTORNEY/AGENT INFORMATION:
NAME: Hoerner Jr., Dennis R.
REGISTRATION NUMBER: 30,914
REFERENCE/DOCKET NUMBER: 38-21(10535)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)537-6099
TELEFAX: (314)537-6047

;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 423 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; PCT-US91-06148A-8

Query Match 9.7%; Score 69; DB 5; Length 423;
Best Local Similarity 24.5%; Pred. No. 5.7;
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

OY 11 KGDDASYEPPTGYLMVNTGVDSKGNELSPHYVEPIK-----PGTTLTKKEIEY 63
DB 190 KGDLVS-----KPYIDITLNLMTKTFGEIENQHQQFVVGKGSYQSPGT-----YL 236
OY 64 VEMALDATAKKEFRVVELDPSAKIEVYYDDKKRKEETKSPITEK 109
DB 237 VEG--DASSASYFLAAAIKGGTVKVTGIGRNSMQGDIRFADVLEK 280

RESULT 48

US-09-243-374-7
Sequence 7, Application US/09243374B
Patent No. 6225114
GENERAL INFORMATION:
APPLICANT: Eichholtz, David Alan
APPLICANT: Gasser, Charles Scott
APPLICANT: Kishore, Ganesh M.
TITLE OF INVENTION: Glycosate-tolerant 5-enolpyruvyl-3-phosphoshikimate
TITLE OF INVENTION: synthases
FILE REFERENCE: EPSP synthases
CURRENT APPLICATION NUMBER: US/09/243,374B
CURRENT FILING DATE: 1999-02-01
EARLIER APPLICATION NUMBER: 07/590,647
EARLIER FILING DATE: 1990-09-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 7
LENGTH: 427
TYPE: PRT
ORGANISM: Escherichia coli
US-09-243-374-7

Query Match 9.7%; Score 69; DB 4; Length 427;
Best Local Similarity 24.5%; Pred. No. 5.7;
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

OY 11 KGDDASYEPPTGYLMVNTGVDSKGNELSPHYVEPIK-----PGTTLTKKEIEY 63
DB 192 KGDLVS-----KPYIDITLNLMTKTFGEIENQHQQFVVGKGSYQSPGT-----YL 238
OY 64 VEMALDATAKKEFRVVELDPSAKIEVYYDDKKRKEETKSPITEK 109
DB 239 VEG--DASSASYFLAAAIKGGTVKVTGIGRNSMQGDIRFADVLEK 282

RESULT 49

US10667-7
Patent No. 5310667
APPLICANT: EICHHOLTZ, DAVID A.; GASSER, CHARLES S.; KISHORE,
GANESH M.
TITLE OF INVENTION: GLYPHOSATE-TOLERANT 5-ENOLPYRUVYL
3-PHOSPHOSHIKIMATE SYNTHASES
NUMBER OF SEQUENCES: 37
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/380,963
FILING DATE: 17-JUL-1989
SEQ ID NO.7:
LENGTH: 427

5310667-7

Query Match 9.7%; Score 69; DB 6; Length 427;
Best Local Similarity 24.5%; Pred. No. 5.7;
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

OY 11 KGDASYEPTGPYLMVNTGVDSKGNELLSPHYVEPIK-----PGTTLTKERIEY 63
DB 192 KGDLYS-----KPYIDITLNMKTFGVEIENHQOOFVVGQSQSPG-----YL 238
OY 64 VEMALDATAKERRVVELDPSAKIEVYYDKNKKKEETKSPITEK 109
DB 239 VEG--DASSASYFLAALAIKGTGVTGIGRNSMGDIRFADYLER 282

RESULT 50

US-08-795-475-3
; Sequence 3, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvick, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMaster, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli LE392/pHDLG, DSM 7055
; US-08-795-475-3

Query Match 9.7%; Score 69; DB 2; Length 434;
Best Local Similarity 31.0%; Pred. No. 5.9;
Matches 36; Conservative 12; Mismatches 48; Indels 20; Gaps 6;

OY 5 DKGYKKGGDASYPEPTGYLMVNTG--VDSKGNELLSPHYVEPIKPGTTLTKERIEY 62
DB 275 ENGKY-----TADLEDGTYTININFACKKYDEKPEEPMDTYKL---ILNKTILKGETTTE 326
OY 63 YVEMALDATAKERRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSE 118
DB 327 AVD---AATALEKVFVKYANDNGVGEWTDYD-----ATKFTYTEKREVI-DASE 372

Search completed: September 1, 2002, 03:18:47
Job time: 4846 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2002, 03:18:49 ; Search time 25.12 Seconds
(without alignments)
209,628 Million cell updates/sec

Title: US-09-728-670-10

Perfect score: 715
Sequence: 1 SSSFDKGRKKKGDASYEP.....SEHNKPGFNLTKVIEKK 136

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.1	99.4	163	1 SAK_STAM	P00802 staphylococ
2	695	97.2	163	1 SAK_BP42	P15240 bacterioph
3	80.5	11.3	3097	1 CADN_DROME	O15943 drosophila
4	78.5	11.0	324	1 YF02_METUA	Q58897 methanococ
5	78.5	11.0	719	1 IF39_TOBAC	P56821 nicotiana t
6	77.5	10.8	442	1 TOP5_BPT4	P07065 bacterioph
7	77	10.8	608	1 PRLR_MOUSE	O08501 mus musculi
8	76.5	10.7	506	1 VIL1_BPV4	P08341 bovine papl
9	75.5	10.6	327	1 LXB1_PROLU	P19840 photorhabdu
10	75	10.5	576	1 NAE2_THEMA	Q9X0Y0 thermotoga
11	74.5	10.4	2511	1 FAS_CHICK	P12276 gallus gall
12	74	10.3	298	1 MOAA_METUA	O58234 methanococ
13	74	10.3	428	1 V234_FOMPV	P14368 fowlpox vit
14	74	10.3	610	1 PRLR_RAT	P05710 rattus norv
15	73.5	10.3	430	1 SERC_ARATH	Q65255 arabidopsis
16	73	10.2	231	1 CTCA_CAUCR	Q45934 caulobacter
17	73	10.2	237	1 TRMD_BUCAI	P57476 buchnera ap
18	73	10.2	263	1 YPJ8_ECOLI	P76612 escherichia
19	73	10.2	2505	1 FAS_RAT	P12785 rattus norv
20	73	10.2	4969	1 RYR2_RABIT	P30957 oryctolagus
21	72.5	10.1	374	1 YEAW_ECOLI	P76253 escherichia
22	72.5	10.1	430	1 SURA_BUCAI	P57240 buchnera ap
23	72.5	10.1	935	1 IF38_MEDTR	Q9X2M1 medicago tr
24	72.5	10.1	1043	1 SVI_CHLPN	O92972 chlamydia p
25	72	10.1	354	1 YO30_BACAN	O92972 chlamydia p
26	72	10.1	429	1 HISX_METUA	Q58881 methanococ
27	72	10.1	695	1 APP2_MOUSE	Q06335 mus musculi
28	72	10.1	978	1 KEMS_RAT	Q06459 rattus norv
29	72	10.1	1102	1 YE20_METUA	O58815 methanococ
30	72	10.1	2663	1 CENE_HUMAN	O02224 homo sapien
31	71.5	10.0	553	1 YF61_METUA	Q58956 methanococ
32	71.5	10.0	560	1 POTR_MYCPN	P75039 mycoplasma
33	71.5	10.0	667	1 Y366_MYCCE	P47606 mycoplasma

34	71.5	10.0	686	1 FREL_YEAST	P32791 saccharomyc
35	71	9.9	344	1 CYSA_STNP7	P14788 synechococ
36	71	9.9	457	1 CRO_DROME	Q27367 drosophila
37	71	9.9	479	1 RFBM_SALTY	P26404 salmoneilla
38	71	9.9	512	1 PUR9_BACSU	P12048 b bifunctio
39	71	9.9	976	1 KEMS_MOUSE	P09581 mus musculi
40	71	9.9	1021	1 YPT7_CAMEL	P41885 caenorhabdi
41	70.5	9.9	534	1 TCEP6_YEAST	P39077 saccharomyc
42	70.5	9.9	559	1 DNLI_PYRKO	Q9HNC4 pyrococcus
43	70.5	9.9	608	1 YD56_YEAST	O04399 saccharomyc
44	70	9.8	115	1 YD46_MYCPN	P75432 mycoplasma
45	70	9.8	251	1 Y255_METUA	O57703 methanococ

ALIGNMENTS

RESULT 1
ID SAK_STAM STANDARD: PRT; 163 AA.
AC P00802;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Staphylokinase precursor (Neutral proteinase) (Protease III).
GN SAK OR SAV1944.
OS Staphylococcus aureus (strain MU50 / ATCC 700699), and
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158878, 1280;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iida T., Ito T.,
Kanamori M., Matsunari H., Maruyama A., Murakami H., Hosoyama A.,
Mizutani-U I., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
Sekiuchi K., Hirakawa H., Kohara S., Goto S., Yabuzaki J.,
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
Hattori M., Ogasawara N., Hayashi H., Hiratake K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
RP SEQUENCE FROM N.A.
RX MEDLINE=84069795; PubMed=6359061;
RA Sako T., Tsuchida N.;
RT "Nucleotide sequence of the staphylokinase gene from Staphylococcus
aureus.";
RL Nucleic Acids Res. 11:7679-7693(1983).
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 43-163.
RX MEDLINE=97290447; PubMed=9145104;
RA Rabinjas A., de Bondt H.L., de Ranter C.;
RT "Three-dimensional structure of staphylokinase, a plasminogen
RT activator with therapeutic potential.";
RL Nat. Struct. Biol. 4:357-360(1997).
[4]
RP STRUCTURE BY NMR OF 28-163.
RX MEDLINE=98367505; PubMed=9692953;
RA Oelenschlaeger O., Ramchandran R., Guehrs K.H., Schlott B.,
Brown L.R.;
RT "Nuclear magnetic resonance solution structure of the plasminogen-
RT activator protein staphylokinase.";
RL Biochemistry 37:10635-10642(1998).
CC -I- FUNCTION: THIS EXTRACELLULAR PROTEIN IS ONE OF THE PLASMINOGEN
CC ACTIVATORS THAT CONVERTS PLASMINOGEN INTO PLASMIN; IT REQUIRES
CC CALCIUM ION FOR STABILIZATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AP003364; BAB58106.1; -
 DR EMBL: X00127; CAA24957.1; -
 DR PIR: A00995; PRSAK.
 DR PDB: 2SAK; 25-FEB-98.
 DR PDB: 1SSN; 02-DEC-98.
 DR InterPro: IPR004093; Staphylokinase.
 DR Pfam: PF02821; Staphylokinase; 1.
 KW Hydrolyase; Calcium; Plasminogen activation; Signal; 3D-structure;
 KM Complete proteome.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 163 STAPHYLOKINASE.
 SO SEQUENCE 163 AA; 18490 MW; E56D9FF50AEDB141 CRC64;

Query Match 99.4%; Score 711; DB 1; Length 163;
 Best Local Similarity 99.3%; Pred. No. 4,1e-59;
 Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SSSFDKGRKKGGDASYPEPTGYLWVNTGYDSKGNELSPHYVEFPPIKPGTTLTKRKI 60
 |||||||
 Db 28 SSSFDKGRKKGGDASYPEPTGYLWVNTGYDSKGNELSPHYVEFPPIKPGTTLTKRKI 87
 Qy 61 EYVEWALDATALKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFPVVDLSEHI 120
 |||||||
 Db 88 EYVEWALDATALKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFPVVDLSEHI 147
 Qy 121 KNPGFNLTTRKVIIEK 136
 |||||||
 Db 148 KNPGFNLTTRKVIIEK 163

RESULT 2
 SAK_BPP42 STANDARD; PRT; 163 AA.
 ID SAK_BPP42
 AC P15240;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Staphylokinase precursor (Neutral protease III).
 GN SAK.
 OS Bacteriophage P42D.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
 OC Lambda phage group.
 OX NCBI_TaxID=10715;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88121731; PubMed=3123893;
 RA Behnke D., Gerlach D.;
 RT "Cloning and expression in *Escherichia coli*. *Bacillus subtilis*, and
 RT *Streptococcus sanguis* of a gene for staphylokinase -- a bacterial
 RT plasminogen activator.";
 RL Mol. Gen. Genet. 210:528-534(1987).
 CC -1- FUNCTION: THIS EXTRACELLULAR PROTEIN IS ONE OF THE PLASMINOGEN
 CC ACTIVATORS THAT CONVERTS PLASMINOGEN INTO PLASMIN; IT REQUIRES
 CC CALCIUM ION FOR STABILIZATION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X06603; CAA29822.1; -
 DR EMBL: M57455; AAA98206.1; -
 DR EMBL: A17537; CAA01341.1; -

DR PIR: S02330; S02330.
 DR HSSP: P00802; 1SSN.
 DR InterPro: IPR004093; Staphylokinase.
 DR Pfam: PF02821; Staphylokinase; 1.
 KW Hydrolyase; Calcium; Plasminogen activation; Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 163 STAPHYLOKINASE.
 SO SEQUENCE 163 AA; 18608 MW; AA267MAFF75D36EC CRC64;

Query Match 97.2%; Score 695; DB 1; Length 163;
 Best Local Similarity 97.8%; Pred. No. 1.2e-57;
 Matches 133; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SSSFDKGRKKGGDASYPEPTGYLWVNTGYDSKGNELSPHYVEFPPIKPGTTLTKRKI 60
 |||||||
 Db 28 SSSFDKGRKKGGDASYPEPTGYLWVNTGYDSKGNELSPHYVEFPPIKPGTTLTKRKI 87
 Qy 61 EYVEWALDATALKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFPVVDLSEHI 120
 |||||||
 Db 88 EYVEWALDATALKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFPVVDLSEHI 147
 Qy 121 KNPGFNLTTRKVIIEK 136
 |||||||
 Db 148 KNPGFNLTTRKVIIEK 163

RESULT 3
 CADN_DROME STANDARD; PRT; 3097 AA.
 ID CADN_DROME
 AC I05943; Q9VJB7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Neurin-cadherin precursor (Cadherin-N protein) (DN-cadherin).
 GN CADN OR CG7100.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Head, and Embryo;
 RX MEDLINE=97388431; PubMed=9247265;
 RA Iwai Y., Usui T., Hirano S., Steward R., Takeichi M., Uemura T.;
 RT "Axon patterning requires DN-cadherin, a novel neuronal adhesion
 RT receptor, in the *Drosophila* embryonic CNS.";
 RL Neuron 19:77-89(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Burton R.C., Rogers J.-H., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Botchan D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Cantor A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertile S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodita C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Meriklo G., Mikhlin N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpston M., Skupski M.P., Smith T.,
 RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Welstock G.M., Weissbach J.,
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RL [3]
 RP INTERACTION WITH ARM.
 RA MEDLINE:98298928; PubMed:9635189;
 RA Loureiro J., Pelletier M.;
 RT "Poles of Armadillo, a Drosophila catenin, during central nervous
 RT system development.";
 RL Curr. Biol. 8:622-632(1998).
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES. MAY ASSOCIATE WITH ARM NEURAL
 CC ISOFORM AND PARTICIPATE IN THE TRANSMISSION OF DEVELOPMENTAL
 CC INFORMATION.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, THE PROTEIN FIRST APPEARS IN
 CC THE MESODERM AT STAGE 9 AND IS PRESENT IN THE MYOBLASTS AND MUSCLE
 CC FIBERS BY STAGE 12 AND STAGE 14, RESPECTIVELY. AT STAGE 12 THE
 CC PROTEIN IS ALSO LOCATED IN THE AXONS OF THE ENTIRE CNS, BUT NOT IN
 CC THE GLIAL CELLS. IN THIRD INSTAR LARVAE PROTEIN IS EXPRESSED IN
 CC THE CNS NEUROPILE, PHOTORECEPTOR AXONS AND PRECURSORS OF ADULT
 CC MUSCLES.
 CC -1- SIMILARITY: CONTAINS 16 CADHERIN DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AB002397; BAA2151.1; -;
 DR EMBL: AE003656; AAF53635.1; -;
 DR HSSP: P00740.11XA.
 DR FlyBase: FBgn0015609; Cadn.
 DR InterPro: IPR002126; Cadherin.
 DR InterPro: IPR000233; Cadherin_C-term.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR001791; Laminin_G.
 DR Pfam: PF00028; cadherin; 14.
 DR Pfam: PF00049; cadherin_C-term; 1.
 DR Pfam: PF00008; EGF; 3.
 DR Pfam: PF00054; laminin_G; 2.
 DR PRINTS: PR00205; CADHERIN.
 DR SMART: SM00112; CA; 16.
 DR SMART: SM00179; EGF; 1.
 DR SMART: SM00001; EGF-like; 3.
 DR SMART: SM00282; Lamg; 2.
 DR PROSITE: PS00232; CADHERIN_1; 9.
 DR PROSITE: PS50268; CADHERIN_2; 16.
 DR PROSITE: PS00022; EGF_1; 3.

DR PROSITE: PS01186; EGF-2; 3.
 DR PROSITE: PS50025; LMV-DOMAIN; 2.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal; EGF-like domain.
 FT SIGNAL 1 36
 FT PROPE 37
 FT CHAIN ? 3097
 FT DOMAIN ? 1454
 FT TRANSMEM 1455 1475
 FT DOMAIN 1476 3097
 FT DOMAIN 181 305
 FT DOMAIN 430 543
 FT DOMAIN 554 651
 FT DOMAIN 660 756
 FT DOMAIN 766 858
 FT DOMAIN 867 968
 FT DOMAIN 978 1078
 FT DOMAIN 1087 1183
 FT DOMAIN 1193 1299
 FT DOMAIN 1307 1414
 FT DOMAIN 1423 1514
 FT DOMAIN 1523 1630
 FT DOMAIN 1639 1742
 FT DOMAIN 1749 1861
 FT DOMAIN 1870 1966
 FT DOMAIN 1974 2085
 FT DOMAIN 2346 2377
 FT DOMAIN 2379 2585
 FT DOMAIN 2592 2627
 FT DOMAIN 2631 2822
 FT DOMAIN 2869 2902
 FT DISULFID 2346 2357
 FT DISULFID 2351 2366
 FT DISULFID 2368 2377
 FT DISULFID 2592 2607
 FT DISULFID 2607 2616
 FT DISULFID 2618 2627
 FT DISULFID 2627 2880
 FT DISULFID 2874 2891
 FT DISULFID 2893 2902
 FT CARBOHYD 97 97
 FT CARBOHYD 150 150
 FT CARBOHYD 325 325
 FT CARBOHYD 426 426
 FT CARBOHYD 930 930
 FT CARBOHYD 1266 1266
 FT VARIANT 1425 1425
 FT CONFLICT 1342 1342
 FT CONFLICT 2786 2786
 SQ SEQUENCE 3097 AA; 347201 MW; 082242F28D9B5CC3 CRC64;
 E -> K (IN ALLELE CADN-M12; MUSCLE
 DEFECTS).
 P -> A (IN REF. 1).
 S -> T (IN REF. 1).
 Query Match 11.3%; Score 80.5; DB 1; Length 3097;
 Best Local Similarity 27.6%; Pred. No. 35;
 Matches 45; Conservative 16; Mismatches 55; Indels 47; Gaps 9;
 QY 13 DASAEPTPEGPYLVNVTGVDSDGNELSP-----HYVEPIKRGTTIKK----- 59
 DB 731 EDSGGFS-TSYDLTIRVTVDVNAAPKFEPLDYQAHVVDSDIPLGYSILVRKAMSDSGSN 789
 QY 60 --TEYVV-----EWALDA-----TAYKEFRVVELD-----PSAKIEVYYDK 94
 DB 790 AEIEIYVSDDHFAVDNGLIIVNNKQJADONNNAIYFIIYATADKGEPPSSVAIYRYTK 849
 QY 95 NKKKEETKSEPTTEKGFVVPDLSEHIKNPGFN-LITKVIIEKK 136
 DB 850 NKNDEE-----PKFSQGVYTPNVD-----NAGPNTLYTVTVASDK 885
 RESULT 4
 YF02_METUA
 ID YF02_METUA STANDARD; PRT; 324 AA.

AC Q56897; DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1502.
GN MJ1502.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
CC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=8688087;
BA Butt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayson R.A., Goeyne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weissflog K.G., Merrick J.M., Glöckner
RA Scott J.L., Geoghegan N.S.M., Weidman J.E., Fuhrman J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,
RA Blank H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii".
SC Science 273:1058-1073(1996).
CC -I- SIMILARITY: BELONGS TO THE ATSA / ELAC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL; U67591; AAB99515.1; -.
DR TIGR; MJ1502; -.
DR InterPro: IPR001279; Beta_lactam_met.
DR Pfam: PF00753; lactamase_B_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 324 AA; 36893 MW; ASD6E2EE3BAAEFM4 CRC64;

Query Match 11.0%; Score 78.5; DB 1; Length 324;
 Best Local Similarity 24.2%; Pred. No.3.6;
 Matches 39; Conservative 18; Mismatches 57; Indels 47; Gaps 7;

QY	7	GKKKGGDASYFEPTGPKPLMVNVTVSDSKGMELDSPHYVEPIPKGTTLTKKEIKIEYYE	65
	: :		
Dd	87	GFEGREKETLKIPGPGTKEIIENSLKLGTHTYEPIKIVEITKEPIITYKE	138
	:	:	
Qy	66	WALDDATAV-----KEFRVELD--PSAKIEVTYIDRKKK-----	97
	:	: :	:
Dd	139	ENEVLIAAPTEHGIPSVAYIFEKIIRKPRDIERAKKLGVKGDPDLKLKNGEAANKIYGE	198
Qy	98	--KEE-----TKSEFTPEKGFPVV--DLSEHIKNPCFNLI	128
Dd	199	IKKPEYVLLPPKKGCFLAYSGDTPLEDFGFKLKLKGCDYL	239

RESULT 5
 IF39_TOBAC STANDARD PRF: 719 AA.

ID IF39_TOBAC AC P56821;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Eukaryotic translation initiation factor 3 subunit 9 (eIF-3 eta) (EIF3 p110).
 GN PTI1.
 OS Nicotiana tabacum (Common tobacco).
 OC Eudicotyledons; Magnoliophyta; Streptocarpyta; Tracheophyta;
 CC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;

```

OC Asteridae: eunasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RC STRAIN=CV, BRIGHT YELLOW 2;
RA Shen W.H., Gigot C.;
RT "Characterization of Ptt1, a gene encoding for one of the subunits of
RL the translation initiation factor 3 (eif3), from Nicotiana tabacum.";
CC Plant Sci. 143:45-54(1999).
CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA.
CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL, Y11996, CAA72721.1; -.
DR InterPro, IPR000504; RRM.
DR Pfam, PF00076; rrm; 1.
DR SMART, SM00360; RRM; 1.
DR PROSITE, PS50102; RRM; 1.
DR PROSITE, PS00030; RRM_RNP_1; FALSE NEG.
KW Initiation factor; Protein biosynthesis; RNA-binding.
FT DOMAIN 60 147 RRM-BINDING (RRM).
SQ SEQUENCE 719 AA; 82524 MW; BBA223B7DA52BB2 CRC64;

Query Match 11.0%; Score 78.5; DB 1; Length 719;
Best Local Similarity 26.2%; Pred. No. 9.4;
Matches 33; Conservative 23; Mismatches 41; Indels 29; Gaps 7;

OY 4 FDKGKYYKKGGDASV-----EPGPIYLVNVTGVDKGN-----ELLSPH 43
DB 543 FDVDELFTMASAEHFMADVDEWDPGRVYATSVTSVHEMENGFNIMSPGKLLYRILNDH 602
OY 44 YVEF--PIKPPETTLTKKEIEYVEMALDAPAY-SEPRVLEDPESAKIEVTYDKNKK-K 98
DB 603 PFQYIMRP-RPPSFLSKERE--ELAKNLRKYSKRYEAEDQDVLSQSDREKRKKLK 658
OY 99 EETKSF 104
DB 659 EEWEMAM 664

RESULT 6
TOP3_BP14 STANDARD; PRT; 442 AA.
AC AC
AC P07065;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA topoisomerase medium subunit (EC 5.9.9.1.3) (Protein Gp52).
GN 52.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
RN [1]
RC SEQUENCE FROM N.A.
RA MEDLINE=87016377; PubMed=3020513;
RA Huang W.M.;
RT "The 52-protein subunit of T4 DNA topoisomerase is homologous to the
RL gyra-protein of gyrase.";
RL Nucleic Acids Res. 14:7379-7390(1986).
RN [2]
RP REVISIONS.

```

RA Huang W.M.:
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kuter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
 RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.,
 RT "Bacteriophage T4 genome analysis."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-46 FROM N.A.
 RX MEDLINE=88172481; PubMed=3280805;
 RA Chapman D., Mord I., Kaufmann G., Galt M.J., Jorissen L., Snyder L.,
 RT "Nucleotide and deduced amino acid sequence of stp: the bacteriophage
 RT T4 anticodon nuclease gene."
 RL J. Mol. Biol. 199:373-377(1998).
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. T4 TOPOISOMERASE
 CC MAKES DOUBLE-STRAND BREAKS.
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
 CC of double-stranded DNA.
 CC -1- SUBUNIT: THIS IS ONE OF THE 3 SUBUNITS OF T4-DNA TOPOISOMERASE.
 CC -1- SIMILARITY: TO THE PROKARYOTIC GYRASE SUBUNIT A AND TO EUKARYOTIC
 CC TOPOISOMERASE II.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04376; CAA27959.1; .
 DR EMBL; AF158101; AAD42487.1; .
 DR PIR; B24705; ITBPT4.
 DR PIR; S01872; S01872.
 DR HSSP; P09097; IAB4.
 DR InterPro: IPR002205; DNA_topoisomIV.
 DR Pfam; PF00521; DNA_topoisomIV; 1.
 DR SMART; SM00434; TOP4C; 1.
 KW Isomerase; Topoisomerase; DNA-binding.
 FT ACT_SITE 117 117 5'-ENDS OF THE DNA VIA A PHOSPHO-TYROSYL
 FT LINKAGE (POTENTIAL).
 FT SEQUENCE 442 AA; 50493 MW; 519E60AE6F5AE7 CRC64;
 SQ
 Query Match 10.88; Score 77.5; DB 1; Length 442;
 Best Local Similarity 25.58; Pred. No. 6.5;
 Matches 36; Conservative 21; Mismatches 43; Indels 41; Gaps 8;
 Oy 16 SYFPTGCVILMVN-VTGVDSKGNELSPHYEFPK-----PGTTLKREKLEYVEWAL 68
 Db 148 AFYLIPTVLNGSGATGATATILHSHSVSKAVIALQSGIKYKPKVE----- 200
 Oy 69 DATAKERR--VEIDPSAKLEVTY-----YKNKKKEETKSF-PITERGF 111
 Db 201 ---FPERGEVEVDGYEIRGTYKFTSRQMTTEIPKYDRETVYSKILDPLENGKF 256
 Oy 112 VVPD--LSEHTKNPGENLTK 130
 Db 257 ITWDACGSH---GFGFRVK 273
 RESULT 7
 PRLR_MOUSE STANDARD: PRT: 608 AA.
 AC 008501; Q62099; P15213; P15212;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Prolactin receptor precursor (PRL-R).
 GN PRLR.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (FORM PRL-R3).
 RC STRAIN=C3H; TISSUE=Mammary gland;
 RX MEDLINE=94085788; PubMed=8262385;
 RA Moore R.C., Oka T.;
 RT "Cloning and sequencing of the cDNA encoding the murine mammary gland
 RT long-form prolactin receptor."
 RL Gene 134:263-265(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (FORM PRL-R3).
 RC STRAIN=SWISS WEBSTER; TISSUE=Liver;
 RX MEDLINE=93307149; PubMed=8319571;
 RA Clarke D.L., Linzer D.I.H.;
 RT "Changes in prolactin receptor expression during pregnancy in the
 RT mouse ovary."
 RL Endocrinology 133:224-232(1993).
 RN [3]
 RP SEQUENCE FROM N.A. (FORM PRL-R3).
 RA Sasaki M.;
 RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. (FORM PRL-R3).
 RC STRAIN=BALB/C; TISSUE=Mammary gland;
 RA Edery M., Pezet A., Nandi S., Kelly P.A.;
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A. (FORMS PRL-R2 AND PRL-R1).
 RC STRAIN=SWISS WEBSTER; TISSUE=Liver;
 RX MEDLINE=89261824; PubMed=2725531;
 RA Davis J.A., Linzer D.I.H.;
 RT "Expression of multiple forms of the prolactin receptor in mouse
 RT liver."
 RL Mol. Endocrinol. 3:674-680(1989).
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR THE ANTERIOR PITUITARY HORMONE
 CC PROLACTIN, AS WELL AS PLACENTAL LACTOGEN I AND II.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: PRL-R1, PRL-R2 AND PRL-R3 (SHOWN
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L13593; AAC37641.1; .
 DR EMBL; L14811; AAA02686.1; .
 DR EMBL; D10214; BAA01066.1; .
 DR EMBL; X73372; CAA51789.1; .
 DR EMBL; M22959; AAA39977.1; .
 DR EMBL; M22958; AAA39976.1; .
 DR PIR; JT0671; JT0671.
 DR HSSP; P14787; IAN3.
 DR MGD; MGI:97763; PRLr.
 DR InterPro: IPR002996; PRLr.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003528; Hematopo_receptor_L_FL.
 DR Pfam; PF00041; fn3; 2.
 DR SMART; SM00060; FN3; 1.
 DR PROSITE; PS01352; HEMATOPO_REC_L_FL; 1.
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat;
 KW Alternative splicing.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 608 PROLACTIN RECEPTOR.
 FT DOMAIN 20 229 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 230 253 BY SIMILARITY.

```

FT DOMAIN 254 608 CYTOPLASMIC (BY SIMILARITY).
FT DOMAIN 20 117 FIBRONECTIN TYPE-III 1.
FT DOMAIN 119 222 FIBRONECTIN TYPE-III 2.
FT DISULFID 31 41 BY SIMILARITY.
FT DISULFID 70 81 BY SIMILARITY.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 281 292 KGSSELSALG -> VANKQLENNY (IN ISOFORM PRL-R2).
FT VARSPLIC 293 608 MISSING (IN ISOFORM PRL-R2).
FT VARSPLIC 281 303 KGSSELSALGCODEPTSC -> IMCSILQLSIVKI
FT VARSPLIC 304 608 PTERICDL (IN ISOFORM PRL-R).
FT CONFLICT 558 558 MISSING (IN ISOFORM PRL-R).
SQ SEQUENCE 608 AA; 68240 MW; B8CE202B2EFC9FC6 CRC64;

Query Match 10.8%; Score 77; DB 1; Length 608;
Best Local Similarity 23.9%; Pred. No. 11;
Matches 34; Conservative 17; Mismatches 43; Indels 48; Gaps 7;

QY 4 FDKGKKKGDASY-----FEPTGPY--LMVNVTVGSDSGNELLSP-HYVEPP---I 49
DB 274 FTHHLEKKSSELLSALGCODEPTSCEDLVLEFLEVDNEDERLMPSHSKETFGQGV 333
QY 50 KRGTVLTKETIEYVWALDATALYKFRVVELDPSAKIEVTYDK---NKKKEETKSP 105
DB 334 KP-----THDDPSDSGHSYDSHSLSEKCEBPQAYP 366
QY 106 ITEKGFVVDLSHINKPGENL 127
DB 367 ---PAHPEITEKPEENPEANI 365

RESULT 8
VIL_BPV4 STANDARD; PRT; 506 AA.
AC P08341;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major capsid protein L1.
GN L1.
OS Bovine papillomavirus type 4.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87282264; PubMed=3039043;
RA Patel K.R., Smith K.T., Campo M.S.;
RT "The nucleotide sequence and genome organization of bovine
papillomavirus type 4."
RT J. Gen. Virol. 68:2117-2128(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X05817; ; NOT_ANNOTATED_CDS.
DR EMBL; D00146; BAA00101.1; -.
DR PIR; B26214; P1WLB4.
DR InterPro; IPR002210; PV_capsid_L1.
DR Pfam; PF00500; late_protein_L1.
DR PRINTS; PR00865; HPVcapsidL1.
DR PRODOM; PD000544; PV_capsid_L1; 1.
KW Coat protein; Late protein.
SQ SEQUENCE 506 AA; 57902 MW; F1C5DFDB54FA681E CRC64;

```

```

Query Match 10.7%; Score 76.5; DB 1; Length 506;
Best Local Similarity 28.7%; Pred. No. 9.5;
Matches 29; Conservative 19; Mismatches 36; Indels 17; Gaps 7;

QY 29 VTGVDG-KG-NELSPHYVEPPKGTTLTKETIEYV---EMALDATALYKFRVVELD 82
DB 331 VTAVDSIRGNFISIVHTPTDEVPQETATYATKHYLRVEM--DLSIMQCLIVNLT 388
QY 83 PSAKIEVTVYDKNKKKEETKSPTEKGFVP--DLSEHIK 121
DB 389 P-----ESTAVLHNNSEITENML---GFIQPPNDIEDHR 422

RESULT 9
LXB1_PROLU STANDARD; PRT; 327 AA.
ID LXB1_PROLU
AC P19840;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Alkanal monooxygenase beta chain (BC 1.14.14.3) (Bacterial luciferase
beta chain).
GN LUXB.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29999;
RX MEDLINE=90375532; PubMed=2204626;
RA Saltner R., Meighen E.;
RT "Nucleotide sequence, expression, and properties of luciferase coded
RT by lux genes from a terrestrial bacterium."
RT J. Biol. Chem. 265:16581-16587(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29999;
RX MEDLINE=92355513; PubMed=1644764;
RA Meighen E.A., Saltner R.B.;
RT "Multiple repetitive elements and organization of the lux operons of
RT luminescent terrestrial bacteria."
RT J. Bacteriol. 174:5371-5381(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HM;
RX MEDLINE=90343746; PubMed=2383248;
RA Johnston T.C., Rucker E.B., Cochrum L., Hruska K.S., Vandegrift V.;
RT "The nucleotide sequence of the luxA and luxB genes of Xenorhabdus
RT luminescens HM and a comparison of the amino acid sequences of
RT luciferases from four species of bioluminescent bacteria."
RT Biochem. Biophys. Res. Commun. 170:407-415(1990).
CC -----
CC -I- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA. THE
CC SPECIFIC ROLE OF THE BETA SUBUNIT IS UNKNOWN, BUT IT IS ABSOLUTELY
CC REQUIRED FOR BIOLUMINESCENCE ACTIVITY.
CC -I- CATALYTIC ACTIVITY: RCHO + FMN(2) + O(2) = RCOOH + FMN + H(2)O +
CC light.
CC -----
CC -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M57416; AAA27624.1; -.
DR EMBL; M90093; AAA27620.1; -.
DR EMBL; M55977; AAA27627.1; -.
DR PIR; C37898; C37898.

```

DR HSSP, P07739, 1BSL.
 DR InterPro: IPR002103; Bac_luciferase.
 DR Pfam: PF00296; bac_luciferase: 1.
 DR PRINTS: PR00089; LUCIFERASE.
 DR PROSITE: PS00494; BACTERIAL_LUCIFERASE: 1.
 KW Photoprotein; Luminescence; Oxidoreductase; Monooxygenase;
 FT Flavoprotein; FMN.
 FT VARIANT 85 85 A -> R (IN STRAIN HM).
 SQ SEQUENCE 327 AA; 37595 MW; 1905AF01658BE56E CRC64;

Query Match 10.6%; Score 75.5; DB 1; Length 327;
 Best Local Similarity 25.4%; Pred. No. 7;
 Matches 33; Conservative 18; Mismatches 44; Indels 35; Gaps 7;

QY 18 FEPGPVLMVAVTGVDSGNELLSPHYVEFPKPGTTL-----TEKIEYVENALDA 70
 DB 163 YTPGGPKRYVAT-----SHHIVEMAKKGIPIEFWMDSNDVREYARAYAVA 212
 QY 71 TAYKEFVVELDPSAKIEVTVYDKNKKKEETKSPITEKGFVVPDLSEHI---KNPGF 125
 DB 213 DKY-DVDSLEIDHQLMLVNNEDSNKAKOETRAF-----ISDYLEHHPNENF 260
 QY 126 -NLTKVIV 134
 DB 261 ENKLEELIAE 270

RESULT 10

NAE2_THEMEA STANDARD; PRT; 576 AA.
 AC Q9X0Y0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable glutamine-dependent NAD(+) synthetase (EC 6.3.5.1) (NAD(+)
 synthase [glutamine-hydrolysing]).
 GN MADE2 OR TM1253.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 CX NCBI_TaxID=2336;

RA [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 Stewart A.M., Colton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).

CC -1- FUNCTION: CAN USE BOTH GLUTAMINE OR AMMONIA AS A NITROGEN
 CC SOURCE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + deamido-NAD(+) + L-glutamine + H(2)O =
 CC AMP + diphosphate + NAD(+) + L-glutamate.
 CC -1- PATHWAY: DE NOVO BIOSYNTHESIS OF NAD.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO THE NAD
 CC SYNTHETASE FAMILY.

CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE001780; AAD36328.1; -
 DR TIGR: TM1253; -
 DR InterPro: IPR003010; CN_hydrolase.

DR InterPro: IPR003694; NAD_synthase.
 DR InterPro: IPR000132; Nitril_cyn_hydrolase.
 DR Pfam: PF00795; CN_hydrolase: 1.
 DR PRINTS: PF02540; NAD_synthase: 1.
 DR PROSITE: PS00920; NITRIL_CHT_1; UNKNOWN_1.
 KW Ligase; NAD; ATP-binding; Complete proteome.
 FT DOMAIN 292 576
 FT NE_BIND 321 328
 FT ACT_SITE 323 323
 SQ SEQUENCE 576 AA; 64919 MW; 208BDC77964C957F CRC64;

Query Match 10.5%; Score 75; DB 1; Length 576;
 Best Local Similarity 18.1%; Pred. No. 15;
 Matches 39; Conservative 31; Mismatches 54; Indels 92; Gaps 7;

QY 7 GKTK-----GDDASFEPTGYLMNV-----TGV 32
 DB 108 GYVRKISLPYGVFDERRYKPGDELLVYKIKVGTICEDINMPVPSASLSGEGV 167
 QY 33 DSKGNELLSPHYVEFPKPGTTLTKKIEYVENA-----IDATA- 72
 DB 168 HLINLSASPHYHGKPYLRKROYLSMKAYDHYAMAYCNMGCODELVFDGSKYVDASGE 227
 QY 73 -----YKEFVVELDPSAKIEVTVYDKNKKKEETKSPF--TEKG----- 110
 DB 228 VYNGKLFEEIITVDLDLENLRVSLYDPRRRYKMTQNPVATVAGNLREKSGHFEPV 287
 QY 111 -----FVVPDLSEHIKNPGFNILTKVI 133
 DB 288 VNPLPVREEMFRALITGLDVRKNGFE--KVI 320

RESULT 11

FAS_CHICK STANDARD; PRT; 2511 AA.
 AC P12276;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
 DE EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
 GN FASN OR FAS.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 CX NCBI_TaxID=9031;

RA [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
 RC STRAIN=WHITE LEGHORN; TISSUE=Liver;
 RX MEDLINE=95031085; PubMed=7944406;
 RA Huang W.-Y., Chitrala S.S., Wakil S.J.;
 RT "Amino-terminal blocking group and sequence of the animal fatty acid
 RT synthase.";
 RL Arch. Biochem. Biophys. 314:45-49(1994).

CC [2]
 CC SEQUENCE OF 75-1775 FROM N.A.
 CC TISSUE=Liver;
 CC MEDLINE=89282777; PubMed=2734291;
 CC Holzer K.P., Liu W., Hammes G.G.;
 CC "Molecular cloning and sequencing of chicken liver fatty acid
 CC synthase cDNA.";
 CC Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).

CC [3]
 CC SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.
 CC MEDLINE=8919426; PubMed=2917973;
 CC Chitrala S.S., Kasturi R., Pazirandeh M., Stojlov D.T., Huang W.-Y.,
 CC Wakil S.J.;
 CC "A novel cDNA extension procedure. Isolation of chicken fatty acid
 CC synthase cDNA clones.";
 CC J. Biol. Chem. 264:3750-3757(1989).
 RA [4]

RP SEQUENCE OF 1752-2512 FROM N.A.
 RX MEDLINE-88320436; PubMed-2842766;
 RA Yuan Z., Liu W., Hammes G.G.;
 RT "Molecular cloning and sequencing of DNA complementary to chicken
 RT liver fatty acid synthase mRNA.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
 RN [5]
 RP SEQUENCE OF 2202-2512 FROM N.A.
 RX MEDLINE-89088152; PubMed-3207710;
 RA Kasuri R., Chirala S.S., Pazirandeh M., Wakil S.J.;
 RT "Characterization of a genomic and cDNA clone coding for the
 RT thioesterase domain and 3' noncoding region of the chicken liver
 RT fatty acid synthase gene.";
 RL Biochemistry 27:7776-7785(1988).
 RN [6]
 RP SEQUENCE OF 2121-2209.
 RX MEDLINE-89192401; PubMed-2648999;
 RA Huang W.-Y., Stoops J.K., Wakil S.J.;
 RT "Complete amino acid sequence of chicken liver acyl carrier protein
 RT derived from the fatty acid synthase.";
 RL Arch. Biochem. Biophys. 270:92-98(1989).
 RN [7]
 RP SEQUENCE OF 2209-2508.
 RX STRAIN-WHITE LEGHORN;
 RT MEDLINE-89088151; PubMed-3207709;
 RA Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
 RT "Complete amino acid sequence of the thioesterase domain of chicken
 RT liver fatty acid synthase.";
 RL Biochemistry 27:7773-7777(1988).
 RN [8]
 RP SEQUENCE OF 667-674 AND 1698-1709.
 RX MEDLINE-89323081; PubMed-2751995;
 RA Chang S.-I., Hammes G.G.;
 RT "Amino acid sequences of pyridoxal 5'-phosphate binding sites and
 RT fluorescence resonance energy transfer in chicken liver fatty acid
 RT synthase.";
 RL Biochemistry 28:3781-3788(1989).
 CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
 CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
 CC THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
 CC ACYL CARRIER PROTEIN.
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
 CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
 CC acetyl-[acyl-carrier protein].
 CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
 CC malonyl-[acyl-carrier protein].
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
 CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
 CC [acyl-carrier protein].
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
 CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
 CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
 CC 2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
 CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
 CC carrier protein] + oleate.
 CC -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
 CC -1- CAUTION: REF. 3 AND REF. 5 SEQUENCES DIFFER FROM THAT SHOWN FROM
 CC POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL: J04485; AAA46389.1; -
 CC DR EMBL: J03860; AAA48767.1; -
 CC DR EMBL: J02839; AAA82106.1; ALT_SEQ.

DR PIR: A33918; XTCFPA.
 DR PIR: A32015; A32015.
 DR HSSP: P02901; IACP.
 DR InterPro: IPR001227; Acyltransf_domain.
 DR InterPro: IPR002085; Adh_zn_family.
 DR InterPro: IPR000794; Ketocacyl-synt.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR InterPro: IPR001031; Thioesterase.
 DR Pfam: PF00698; Acyl_transf_1.
 DR Pfam: PF00107; adh_zinc_1.
 DR Pfam: PF00109; ketocacyl-synt_1.
 DR Pfam: PF02801; ketocacyl-synt_C_1.
 DR Pfam: PF00550; pp-binding_1.
 DR Pfam: PF00975; Thioesterase_1.
 DR ProSite: PS00012; PHOSPHOPANTETHEINE_1.
 DR ProSite: PS00606; B_KETOACYL_SYNTHASE_1.
 DR ProSite: PS50075; ACP_DOMAIN_1.
 KW Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
 KW Transferrase; Hydrolase; Oxidoreductase; lyase; NADP; acetylation;
 KW Alternative splicing; Pyridoxal phosphate.
 FT INT_MET 0
 FT DOMAIN 1 7411 0
 FT BINDING 427 815
 FT DOMAIN 1638 1866
 FT DOMAIN 1867 2119
 FT DOMAIN 2124 2180
 FT DOMAIN 2209 2511
 FT MOD_RES 1
 FT ACT_SITE 160 160
 FT ACT_SITE 579 579
 FT ACT_SITE 877 877
 FT NP_BIND 1674 1691
 FT BINDING 1707 1707
 FT NP_BIND 1888 1903
 FT BINDING 2157 2157
 FT ACT_SITE 2308 2308
 FT ACT_SITE 2481 2481
 FT VARSPIC 2348 2348
 FT CONFLICT 77 78
 FT CONFLICT 116 116
 FT CONFLICT 675 675
 FT CONFLICT 1169 1169
 FT CONFLICT 1178 1178
 FT CONFLICT 1191 1191
 FT CONFLICT 1198 1198
 FT CONFLICT 1286 1287
 FT CONFLICT 1372 1372
 FT CONFLICT 1533 1533
 FT CONFLICT 1577 1577
 FT CONFLICT 1685 1696
 FT CONFLICT 1732 1732
 FT CONFLICT 1745 1745
 FT SEQUENCE 2511 AA; 274648 MW; 6220393DAC8315D3F C6C64;
 SO

Query Match 10.4%; Score 74.5; DB 1; Length 2511;
 Best Local Similarity 25.5%; Pred. No. 98;
 Matches 26; Conservative 13; Mismatches 24; Indels 39; Gaps 5;
 Oy 27 VNTGVDSKGNELSPHYVEPIKPGTTLTKETIEYVEMALDPAVKKEFRVELDPSAK 86
 Db 801 IHLGINVGNLPP-VEYVPVGPPL-----ISPIKWK-----DHSQD 839
 Oy 87 IEVTVYDKNKKKEKSPITKEGF-----VVDLSH 119
 Db 840 WDV-----PRADEPSSGKSASASYNDIVSPDSPH 872

RESULT 12
 MOAA_METUA

```

ID MOA_METUA STANDARD; PRT; 298 AA.
AC 058234;
AD 01-NOV-1997 (Rel. 35, Created)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative molybdopter in cofactor synthesis protein A.
GN MOA OR M00824.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=6888087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.T.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Fritch J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhlmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.;
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF MOLYBDOPTERIN PRECURSOR
CC 2 FROM GUANOSINE (BY SIMILARITY).
CC -1- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MOA / NIFB / POOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67526; AAB98823.1; -
DR InterPro: IPR000385; MOA_NiFe_Pqqe.
DR Pfam; PF04444; MOA_NiFe_Pqqe.1.
DR PROSITE; PS01305; MOA_NiFe_POOE.1.
KM Molybdenum cofactor biosynthesis; Iron-sulfur; Complete proteome.
FT METAL 20 20 IRON-SULFUR (POTENTIAL).
FT METAL 24 24 IRON-SULFUR (POTENTIAL).
FT METAL 27 27 IRON-SULFUR (POTENTIAL).
FT SEQUENCE 298 AA; 34652 MW; EA92B9A0A75BED55 CRC64;
SO

```

```

Query Match 10.3%; Score 74; DB 1; Length 298;
Best Local Similarity 21.4%; Pred. No. 8.6;
Matches 21; Conservative 19; Mismatches 42; Indels 16; Gaps 1;

```

```

OY 31 GVDKGNLSPHYEPPIKPGCTTLTKREIYVEMALDATAYKFRVVELDPSAKIEVT 90
DB 31 GHDSNNDYMPPEELGIAKTSTKEGVAKI-----KISGEPLLRKQVC 74
OY 91 YVDKKKKEKSPITEKGFVVDLSHINKNPGFNLA 128
DB 75 ELENIKDERIKDISLTJNGILLELAKELKDAGLNRY 112

```

RESULT 13

```

V234_F0MPV STANDARD; PRT; 428 AA.
ID V234_F0MPV
AC P14368; P14367; Q9J501;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ankrylin-repeat protein FPV234 (BamHI-ORF12/ORF13).
GN FPV234.

```

```

OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus."
RL J. Virol. 74:3815-3831(2000).
RN [2]
RP SEQUENCE OF 65-428 FROM N.A.
RC STRAIN=FP-9 / Isolate HP-438;
RX MEDLINE=88229622; PubMed=2836548;
RA Tomlay F., Binnis M., Campbell J., Boursnell M.E.G.;
RT "Sequence analysis of an 11.2 kilobase, near-terminal, BamHI fragment
RT of fowlpox virus."
RL J. Gen. Virol. 69:1025-1040(1988).
CC -1- SIMILARITY: CONTAINS 8 ANK REPEATS.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in position 204, 219 and 237.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF198100; AAF44578.1; -
DR EMBL; D00295; BAA00209.1; ALT_FRAME.
DR EMBL; D00295; BAA00207.1; ALT_FRAME.
DR PIR; C30087; KMWZTW.
DR PIR; D30087; KMWZTW.
DR InterPro: IPR002110; ANK.
DR Pfam; PF00003; ank; 4.
DR SMART; SM00248; ANK_REPEAT; 5.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KM Hypothetical protein: Repeat; ANK repeat.
FT REPEAT 6 35 ANK 1.
FT REPEAT 39 68 ANK 2.
FT REPEAT 71 100 ANK 3.
FT REPEAT 103 132 ANK 4.
FT REPEAT 137 169 ANK 5.
FT REPEAT 174 202 ANK 6.
FT REPEAT 206 238 ANK 7.
FT REPEAT 242 271 ANK 8.
FT SEQUENCE 428 AA; 49194 MW; C71BEAF74687EB8F CRC64;
SO

```

```

Query Match 10.3%; Score 74; DB 1; Length 428;
Best Local Similarity 26.6%; Pred. No. 13;
Matches 29; Conservative 14; Mismatches 38; Indels 28; Gaps 4;

```

```

OY 43 HYVEPIKPGCTTLTKERI-EYVEMALDATA-----YKEFRV---VEL 81
DB 102 HYCKPFIHIANRTESTIVALLIEYGADINSEGANGKYPYHYAMKVDYDFRKAITIKVLL 161
OY 82 DPSAKIEVTYYDKNNKKKEKSPITEKGFVVDLSHINKNPGFNLA 130
DB 162 DHGA-----DINKOSVLTNTSPLEYETRTDLDLYISRGANINIK 203

```

RESULT 14

```

PRLR_RAT STANDARD; PRT; 610 AA.
ID PRLR_RAT
AC P05710; Q63451; Q63723; Q62832; Q64274; Q63479;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Prolactin receptor precursor (PRL-R) (Lactogen receptor).
GN

```

GN PRLR.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91155946; PubMed=2293022;
 RA Shirota M., Banville D., Ali S., Jolicoeur C., Boutin J.M.,
 RA Edery M., Djiane J., Kelly P.A.;
 RT "Expression of two forms of prolactin receptor in rat ovary and
 RT liver";
 RL Mol. Endocrinol. 4:1136-1143(1990).
 RN [2]
 RP SEQUENCE FROM N.A. (LONG FORM AND SHORT FORM).
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Ovary;
 RX MEDLINE=90241201; PubMed=2159291;
 RA Zhang R., Bucko E., Tsai-Morris C.H., Hu Z.Z., Dufau M.L.;
 RT "Isolation and characterization of two novel rat ovarian lactogen
 RT receptor cDNA species";
 RL Biochem. Biophys. Res. Commun. 168:415-422(1990).
 RN [3]
 RP SEQUENCE OF 281-610 FROM N.A.
 RA Banville D., Stocco R., Murthy K.K., Bole Y., Kelly P.A.;
 RL Submitted (Mar-1996) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (MEDIUM FORM).
 RC TISSUE=Liver;
 RX MEDLINE=88165059; PubMed=2832068;
 RA Boutin J.-M., Jolicoeur C., Okamura H., Gagnon J., Edery M.,
 RA Shirota M., Banville D., Dusanter-Fourt I., Djiane J., Kelly P.A.;
 RT "Cloning and expression of the rat prolactin receptor, a member of
 RT the growth hormone/prolactin receptor gene family";
 RL Cell 53:69-77(1988).
 RN [5]
 RP SEQUENCE FROM N.A. (FORM NB2).
 RC TISSUE=Lymphoma;
 RX MEDLINE=92041834; PubMed=1718958;
 RA Ali S., Pelligrini I., Kelly P.A.;
 RT "A prolactin-dependent immune cell line (NB2) expresses a mutant form
 RT of prolactin receptor";
 RL J. Biol. Chem. 266:20110-20117(1991).
 RN [6]
 RP SEQUENCE FROM N.A. (FORM NB2).
 RX MEDLINE=95014332; PubMed=7929319;
 RA O'Neal K.D., Yu-Lee L.Y.;
 RT "Differential signal transduction of the short, NB2, and long
 RT prolactin receptors. Activation of interferon regulatory factor-1 and
 RT cell proliferation";
 RL J. Biol. Chem. 269:26076-26082(1994).
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR THE ANTERIOR PITUITARY HORMONE
 CC PROLACTIN.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS ARE PRODUCED BY ALTERNATIVE
 CC SPLICING OF THE PRLR GENE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M57668; AAA41938.1; -
 DR EMBL: M34083; AAA79273.1; -
 DR EMBL: L48060; AAA9274.1; -
 DR EMBL: U34730; AAA92053.1; -
 DR EMBL: M19304; AAA41937.1; -
 DR EMBL: M74152; AAA41946.1; -
 DR EMBL: U07567; AAA61784.1; -

DR PIR: A29884; A29884.
 DR HSSP: P14787; IAN3.
 DR InterPro: IPR002996; CRIA.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003528; Hematopo_receptor_L_F1.
 DR Pfam: PF00041; fn3; 2.
 DR SMART: SM00060; FN3; 1.
 DR PROSITE: PS01352; HEMATOPO_REC_L_F1; 1.
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat;
 KW Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 610
 FT DOMAIN 20 229
 FT TRANSMEM 230 253
 FT DOMAIN 254 610
 FT DOMAIN 20 117
 FT DOMAIN 119 222
 FT DISULFID 31 41
 FT DISULFID 70 81
 FT CARBOHYD 54 54
 FT CARBOHYD 99 99
 FT CARBOHYD 127 127
 FT VARSPPLIC 131 150
 FT FT 151 610
 FT VARSPPLIC 281 310
 FT FT 311 610
 FT VARSPPLIC 342 539
 FT CONFLICT 236 236
 FT CONFLICT 345 345
 FT CONFLICT 465 465
 FT CONFLICT 466 466
 FT CONFLICT 469 469
 FT CONFLICT 541 541
 FT CONFLICT 555 555
 SQ SEQUENCE 610 AA; 68599 MW; 83D04D832861295D CRC64;
 Query Match 10.3%; Score 74; DB 1; Length 610;
 Best local similarity 23.9%; Pred. No. 20;
 Matches 34; Conservative 17; Mismatches 43; Indels 48; Gaps 7;
 QY 4 FDGKRYKKGDADASY-----EPTGPY--LWNVNTGDSKGNELSP-HYVEFP--I 49
 DB 274 FDTLLEKSGSEELLSALGCDPPPTSDCEDLVEFLVDNDNEDERLMPHSKEYGGGV 333
 QY 50 KPCTTLTKKEIEYVWALDATALKFRVLEDPSSAKIEVTYDK----NKKKEPKSP 105
 DB 334 KP-----THLPDSDSGHSGYDSHSLSEKCEPQAYP 366
 QY 106 ITEKGFVVDLSEHKNPGENL 127
 DB 367 PT---LHIEPTEKPENPENAI 385
 RESULT 15
 SERC ARATH STANDARD; PRT; 430 AA.
 ID SERC ARATH
 AC 096255;
 DT 15-DEC-1998 (rel. 37, Created)
 DT 15-DEC-1998 (rel. 37, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Phosphoserine aminotransferase, chloroplast precursor (EC 2.6.1.52)
 DE (PSAT).
 DE AT4G35630 OR F8D20.140.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 CC NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE-99097839; PubMed-9881164;
 RA Ho C.-L., Noji M., Saito M., Yanagaki M., Saito K.;
 RT "Molecular characterization of plastidic phosphoserine
 aminotransferase in serine biosynthesis from Arabidopsis.";
 RL Plant J. 16:443-452(1998).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLDBIA;
 RX MEDLINE-20083488; PubMed-10617198;
 RA Meyer K.F.X., Schueller C., Mambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Macho R., Mueller M.,
 RA Kreis M., Delseny M., Pulidomench P., Watson M., Schmidtmann T.,
 RA Reichert B., Portelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Blham L., Robben J.,
 RA Van der Schueren J., Gymnopoulos B., Chuang Y.-J., Vandenbussche F.,
 RA Breken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weltzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirks W.,
 RA Moollman S., Klein lankhorst R., Rose M., Haut J., Koetter P.,
 RA Berneriser S., Hempel S., Feldpausch M., Lambert S., Van den Daele H.,
 RA De Keyser A., Blysshaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., Molay K., Mayes R.,
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Blocker H., Scharf M., Grimm M., Loehert T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Hetzl A.,
 RA Neumann S., Argitlon A., Vitale D., Liguori R., Piravandi E.,
 RA Maassen O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabel S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chedfor F., Cooke R., Berger C., Monfort A., Casqueria E.,
 RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijman L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Pannell L., Dedha N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Mix P., Bentley D., Fulton R., Miller N., Greco T., Kemp K.,
 RA Kriemer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spiehl J., Ryan E., Andrews S., Gelsel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Grana S., Shohdy N., Hasegawa A., Hameed A., Loch M., Johnson A.,
 RA Chen E., Maria M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- CATALYTIC ACTIVITY: O-phospho-L-serine + 2-oxoglutarate = 3-
 CC phosphonoxypropylate + L-glutamate.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- PATHWAY: MAJOR PHOSPHORYLATED PATHWAY OF SERINE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Chloroplast.
 CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL; D88541; BAA13640.1; -

DR EMBL; AB010408; BAA24441.1; -
 DR EMBL; AL031135; CAA20033.1; -
 DR EMBL; AL161587; CAB80279.1; -
 DR HSSP; P23721; IBOO.
 DR InterPro; IPR000192; Aminotransf.-class-V
 DR InterPro; IPR003248; Phosphat.-aminotransf.
 DR Pfam; PF00266; aminotran_5; 1.
 DR ProDom; PD001544; Phosphat.-aminotransf.; 1.
 DR PROSITE; PS00595; AA-TRANSFER. CLASS.5; 1.
 DR Serine biosynthesis; Transferase; Aminotransferase;
 DR Pyridoxal phosphate; Chloroplast; Transf. peptide.
 DR TRANSIT 1
 DR CHAIN 1
 DR BINDING 265 265
 DR SEQUENCE 430 AA; 47359 MW; B3225CC8DE78BDE2 CRC64;
 SQ
 Query Match 10.3%; Score 73.5; DB 1; Length 430;
 Best Local Similarity 22.1%; Pred. No. 15;
 Matches 33; Conservative 17; Mismatches 48; Indels 51; Gaps 4;
 QY 11 KGDASYEPTGPLYLWNTGVDSKGNELLSPHYVEFPKPGTTLKEIEY---YVMA 67
 DB 158 KSDPTVDFVYTGSM-----GDKAVKEAKKTKTNVWS 190
 QY 68 LDATAYKE--FRVVELDPSAKIEVYTDKNNKKEETKSPITEKGVVDLSEHI---- 120
 DB 191 GKSEKTKVPSFELBQTPDAKYLHCANETIHGVFEKQYPPKNGFLVADMSNCSKP 250
 QY 121 -----KNPFLNLIKVIIEK 135
 DB 251 VDVSKEGVYIGAOKNVGPGVTVIIRK 279
 RESULT 16
 ID CTRR_CAUOR STANDARD; PRT; 231 AA.
 AC 045994;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE Cell cycle transcriptional regulator crra (response regulator soka).
 GN CTRR OR SOKA OR CC3035.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CH15N / NA1000;
 RX MEDLINE-96140642; PubMed-8548829;
 RA Quon K.C., Marczyński G.T., Shapiro L.;
 RT "Cell cycle control by an essential bacterial two-component signal
 transduction protein.";
 RL Cell 84:83-93(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE-21173698; PubMed-11259647;
 RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Padke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,
 RA Uitterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 RN [3]
 RP SEQUENCE OF 159-231 FROM N.A.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE-98132609; PubMed-9465034;
 RA Wu J., Ohta N., Newton A.;

"An essential, multicomponent signal transduction pathway required for cell cycle regulation in *Caulobacter*.";
 Proc. Natl. Acad. Sci. U.S.A. 95:1443-1448(1998).
 CC -1- FUNCTION: FORMS PART OF A TWO-COMPONENT REGULATORY SYSTEM
 CC CTRN/CCNA THAT CONTROLS MULTIPLE EVENTS IN THE CELL CYCLE,
 CC INCLUDING CELL DIVISION, STALK SYNTHESIS, AND CELL CYCLE-SPECIFIC
 CC TRANSCRIPTION. BINDS TO A GROUP OF CELL CYCLE-REGULATED PROMOTERS
 CC CRITICAL FOR DNA REPLICATION, DNA METHYLATION, AND CLASS II
 CC FLAGELLAR BIOGENESIS.
 CC -1- PTM: PHOSPHORYLATED BY CCKA.
 CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U39559; AAA93080.1; -;
 CC EMBL: AE005966; AAK2497.1; -;
 CC EMBL: AF021339; AAC05479.1; -;
 CC HSSP: P08402; 1B00.
 CC TIGR: CC3035; -;
 CC InterPro: IPR001789; Response_Reg.
 CC InterPro: IPR001867; Trans_Reg.C.
 CC Pfam: PF00072; response_reg.1.
 CC Pfam: PF00486; trans_reg.C.1.
 CC SMART: SM00448; REC.1.
 CC PROSITE: P550110; RESPONSE_REGULATORY; 1.
 CC Sensory transduction; Phosphorylation; Transcription regulation;
 CC DNA-binding; Complete proteome.
 CC FT DOMAIN 1 116 RESPONSE REGULATORY.
 CC MOD_RES 51 51 PHOSPHORYLATION (BY SIMILARITY).
 CC SEQUENCE 231 AA; 25796 MW; B8B12AF6EDADB8 CRC64;
 SQ
 Query Match 10.2%; Score 73; DB 1; Length 231;
 Best Local Similarity 26.2%; Pred. No. 7.9;
 Matches 32; Conservative 19; Mismatches 39; Indels 32; Gaps 5;
 QY 10 KKGD-----DASYEPTGYLWVNGVDSKGNELSPHYVEPIKPGTTLKKEIEYV 64
 DB 128 KTGDIYVNDAKVEVNGN--RVHLTGKEYOMLELS-----LKKGTTLKKEFLNL 178
 QY 65 EWALDNTAYKEFRVVELDPSAEIYVYDKNKKETKSF-----PIKEGFVVDLSE 118
 DB 179 YGMD-----EBELKIDIVFICKLRKKLAASHGKHHTETVWGRGVLRDPRE 226
 QY 119 HI 120
 DB 227 QV 228
 RESULT 17
 TRMD_BUCAT
 ID TRMD_BUCAT STANDARD; PRT; 237 AA.
 AC P57476;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE tRNA (Guanine-N1)-methyltransferase (EC 2.1.1.31) (MIG-
 DE methyltransferase) (tRNA [GM37] methyltransferase).
 GN TRMD OR B0396.
 OS *Buchnera aphidicola* (subsp. *Acyrthosiphon pisum*) (*Acyrthosiphon pisum*
 OS symbiotic bacterium).
 CC Bacteria; Proteobacteria; gamma subdivision; *Buchnera*.
 CC NCB1_TaxID=18099;
 RN NCB1_TaxID=18099;
 RP SEQUENCE FROM N.A.
 RC STRAIN=TORO 1998;
 RX MEDLINE=20445173; PubMed=10993077;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RT *Buchnera* sp. *Apf.*."
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: SPECIFICALLY METHYLATES GUANOSINE-37 IN VARIOUS TRNAS
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
 CC homocysteine + tRNA containing N1-methylguanine.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRMD FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AP001119; BAB13099.1; -;
 CC InterPro: IPR002649; tRNA_mig_MT.
 CC Pfam: PF01746; tRNA_mig_MT.1.
 CC ProDom: PD004978; tRNA_mig_MT.1.
 CC Transferrase; Methyltransferase; tRNA processing; Complete proteome.
 CC SEQUENCE 237 AA; 27335 MW; 86CE383B35EC4AB CRC64;
 SQ
 Query Match 10.2%; Score 73; DB 1; Length 237;
 Best Local Similarity 25.5%; Pred. No. 8.1;
 Matches 36; Conservative 17; Mismatches 40; Indels 48; Gaps 8;
 QY 3 SPDKKTKKGDASY-----PEPTGYLWV-----NVTGV--DSKGNELSPH 43
 DB 49 NFSKNKTVSDRYGGGPGMLMSFEPL--YLAIQAKSTKNTVYILSPQELKQNH 106
 QY 44 YVEPIKPGTTLKKEIEYV-----EWALD--TAYKEFRVVELDP 83
 DB 107 IEELVKK-----KKYIFIGRYEGIDQRIIDNVDEWSIGSYILTGELAAWIDA 160
 QY 84 SAKIEVYVYDKNKKETKSF 104
 DB 161 ISRL-IPGVITKKSIEDSF 180
 RESULT 18
 YPJ_BECOLI
 ID YPJ_BECOLI STANDARD; PRT; 263 AA.
 AC P7612;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ypbJ.
 GN YPJ OR B2649.
 OS *Escherichia coli*.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC *Escherichia*.
 CC NCB1_TaxID=562;
 RN NCB1_TaxID=562;
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1464(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

FT ACT_SITE 161 161 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT ACT_SITE 581 581 MALONYLTRANSFERASE (BY SIMILARITY).
 FT NP_BIND 1662 1682 NADP (ER).
 FT BINDING 1698 1698 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 FT NP_BIND 1765 1780 NADP (KR).
 FT BINDING 2151 2151 PHOSPHOANETHINE (BY SIMILARITY).
 FT ACT_SITE 2302 2302 THIOESTERASE (BY SIMILARITY).
 FT ACT_SITE 2475 2475 THIOESTERASE (BY SIMILARITY).
 FT ACT_SITE 878 878 BETA-HYDROXYACYL DEHYDRATASE (BY SIMILARITY).
 FT CONFLICT 871 871 S -> P (IN REF. 3).
 FT CONFLICT 1967 1968 MY -> IL (IN REF. 5).
 FT CONFLICT 2085 2085 C -> P (IN REF. 4).
 FT CONFLICT 2106 2106 A -> V (IN REF. 1 AND 5).
 FT CONFLICT 2296 2296 Y -> H (IN REF. 1 AND 5).
 SQ SEQUENCE 2505 AA; 272647 MW; 5810EC13D37F3114 CRC64;

Query Match 10.2%; Score 73; DB 1; Length 2505;
 Best Local Similarity 30.0%; Pred. No. 1.3e+02;
 Matches 24; Conservative 14; Mismatches 32; Indels 10; Gaps 4;

OY 27 VAVTGVDSGNELLSPHYVEPIKPTLTKEIEYVEMALDAVAKREYVELDP--S 84
 DB 803 VHTGIDIPNMLFP--VEFFVPRKSTPL-----ISPHNM--DHSQWDIPVAEDFPNGS 854

OY 85 AKIEVTVYDKNKKKEETKSF 104
 DB 855 SSSSATVYNIDASSSESDHY 874

RESULT 20
 RYR2_RABIT STANDARD; PRT; 4969 AA.
 ID RYR2_RABIT
 AC P30957;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ryanodine receptor 2 (Cardiac muscle-type ryanodine receptor) (RYR2)
 DE (RYR-2) (Cardiac muscle ryanodine receptor-calcium release channel).
 GN RYR2.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OC NCBI_TaxID=9986;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart muscle;
 RX MEDLINE=90337947; PubMed=2380170;
 RA Otsu K., Willard H.F., Khanna V.K., Zorzato F., Green N.M.,
 MacLennan D.H.;
 RA "Molecular cloning of cDNA encoding the Ca2+ sarcoplasmic
 RT (ryanodine receptor) of rabbit cardiac muscle sarcoplasmic
 RT reticulum.";
 RT J. Biol. Chem. 265:13472-13483(1990).
 RN [2]
 RP PHOSPHORYLATION OF SER-2809.
 RX MEDLINE=91250425; PubMed=1645727;
 RA Wichter D.R., Kovacs R.J., Schulman H., Cefali D.C., Jones L.R.;
 RT "Unique phosphorylation site on the cardiac ryanodine receptor
 RT regulates calcium channel activity.";
 RT J. Biol. Chem. 266:11144-11152(1991).
 RL
 CC -1- FUNCTION: Communication between transverse-tubules and
 CC sarcoplasmic reticulum. Contraction of cardiac muscle is triggered
 CC by release of calcium ions from SR following depolarization of T-
 CC tubules.
 CC -1- SUBUNIT: Homotrimer (Potential).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: HEART AND BRAIN.
 CC -1- MISCELLANEOUS: The calcium release channel is modulated by calcium
 CC ions, magnesium ions, ATP and calmodulin.
 CC -1- MISCELLANEOUS: The calcium release channel activity resides in the
 CC C-terminal region while the remaining part of the protein

CC constitutes the 'foot' structure spanning the junctional gap
 CC between the SR and the T-tubule. It is possible that the foot
 CC structure interacts with the cytoplasmic region of the
 CC dihydropyridine receptor.
 CC -1- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-
 CC -1- release channel in junctional SR and modulates its activity.
 CC -1- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M59743; AAA31179.1; -.
 DR PIR: A37113; A37113.
 DR InterPro: IPR000636; Cation_chan_non_lig.
 DR InterPro: IPR001682; Channel_pore_Ca_Na.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR003608; MIR.
 DR InterPro: IPR000699; RYR_nodn_receptor.
 DR InterPro: IPR003032; RYR.
 DR InterPro: IPR001215; Ryanodn_receptor.
 DR InterPro: IPR003877; SPRY.
 DR InterPro: IPR003878; SPRY_domain.
 DR Pfam: PF00036; efhand; 2.
 DR Pfam: PF00520; ion.trans; 1.
 DR Pfam: PF02815; MIR; 4.
 DR Pfam: PF01365; RYR_ITPR; 2.
 DR Pfam: PF02026; RYR; 4.
 DR Pfam: PF00622; SPRY; 3.
 DR PRINTS: PRO0795; RYANODINER.
 DR SMART: SM00449; SPRY; 4.
 DR SMART: SM00472; MIR; 4.
 KW Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;
 KW Phosphorylation; Glycoprotein; Multigene family; Calmodulin-binding.
 FT DOMAIN 1 3090
 FT TRANSMEM 3091 3110 M' (POTENTIAL).
 FT TRANSMEM 3154 3172 M'' (POTENTIAL).
 FT TRANSMEM 3941 3960 M1 (POTENTIAL).
 FT TRANSMEM 3979 3996 M2 (POTENTIAL).
 FT TRANSMEM 4234 4257 M3 (POTENTIAL).
 FT TRANSMEM 4295 4315 M4 (POTENTIAL).
 FT TRANSMEM 4501 4521 M5 (POTENTIAL).
 FT TRANSMEM 4580 4602 M6 (POTENTIAL).
 FT TRANSMEM 4722 4742 M7 (POTENTIAL).
 FT TRANSMEM 4770 4788 M8 (POTENTIAL).
 FT TRANSMEM 4812 4829 M9 (POTENTIAL).
 FT TRANSMEM 4847 4869 M10 (POTENTIAL).
 FT DOMAIN 670 808 SPRY 1.
 FT DOMAIN 1098 1221 SPRY 2.
 FT DOMAIN 1423 1561 SPRY 3.
 FT DOMAIN 853 2926 4 X APPROXIMATE REPEATS.
 FT REPEAT 853 966 1.
 FT REPEAT 967 1080 2.
 FT REPEAT 2693 2811 3.
 FT REPEAT 2813 2926 4.
 FT BINDING 2619 3016 MODULATOR (POTENTIAL).
 FT BINDING 2877 2898 CALMODULIN (POTENTIAL).
 FT BINDING 2938 3016 CALMODULIN (POTENTIAL).
 FT MOD_RES 2809 2809 PHOSPHORYLATION (BY CAM-KINASE).
 FT CARBOHYD 4105 4105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4796 4796 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 4969 AA; 565060 MW; FFE0684B974DB4D CRC64;

Query Match 10.2%; Score 73; DB 1; Length 4969;
 Best Local Similarity 25.0%; Pred. No. 3e+02;
 Matches 27; Conservative 20; Mismatches 35; Indels 26; Gaps 6;

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AP001118; BAB12858.1; -
 DR InterPro: IPR000297; Rotamase.
 DR Pfam: PF00639; Rotamase; 1.
 DR PROSITE: PS01096; PPIG_PPIASE_1; FALSE_NEG.
 DR PROSITE: PS50198; PPIG_PPIASE_2; 2.
 KM Isomerase; Rotamase; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 1 430 SURVIVAL PROTEIN SURA HOMOLOG.
 FT DOMAIN 179 277 PPIG 1.
 FT DOMAIN 286 386 PPIG 2.
 SQ SEQUENCE 430 AA; 50737 MW; CAAC425B3FBA5788 CRC64;

Query Match 10.1%; Score 72.5; DB 1; Length 430;
 Best Local Similarity 22.2%; Pred. No. 18;
 Matches 39; Conservative 28; Mismatches 44; Indels 65; Gaps 12;

QY 9 YKGGDASYEPTGPLYL-----MVNVTCVDSKGNELSPHYVER-PIKPGTTLF-----K 57
 Db 253 FKKG-----QIGPIYVGKGLYILKNDIHKKENIVTEFTMOCLIKPSYILINTEAK 306
 QY 58 EKIEYVEWALDATAYKEER--VVELDPSAK-IEVTVYDKNK-----KEETKSEPTTE 108
 Db 307 KRT-----FNIEYENIKKGIYFDDAKNLSDDYSSNKKGLGWSIKESLGFIDLK 357
 QY 109 KGVVVDLSE-----H-----KNGGFLNI--TKVIEK 135
 Db 358 K-FLIDKNEISEPVKSNMGWHIFKILDRQVDAPFYKLLKKNQAFNIVLKNKITSK 412

RESULT 23
 IF38_MEDTR STANDARD; PRT; 935 AA.
 AC 09XHM1;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Probable eukaryotic translation initiation factor 3 subunit 8 (eIF3
 p110).
 GN AM3-1.
 OS Medicago truncatula (Barrel medic).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucroideae; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
 OX NCBI_TaxID=3880;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A17;
 RA MEDLINE-99164949; PubMed-10065555;
 RA van Buren M.L., Maldonado-Mendoza I.E., Tieu A.T., Blylock L.A.,
 Harrison M.J.;
 RT Novel genes induced during an arbuscular mycorrhizal (AM) symbiosis
 formed between Medicago truncatula and Glomus versiforme.";
 RT Mol. Plant Microbe Interact. 12:171-181(1999).
 RL -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
 METHIONYL-TRNAI AND MRNA (BY SIMILARITY).
 CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE EIF38 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF106930; AAD39891.1; -
 DR InterPro: IPR000717; PCI.
 DR Pfam: PF01399; PCI; 1.
 DR SMART: SM00088; PINTW; 1.
 KM Initiation factor; Protein biosynthesis.
 SQ SEQUENCE 935 AA; 105069 MW; 199F5FF898EB11F CRC64;

Query Match 10.1%; Score 72.5; DB 1; Length 935;
 Best Local Similarity 23.0%; Pred. No. 46;
 Matches 37; Conservative 31; Mismatches 50; Indels 43; Gaps 11;

QY 1 SSSFDKGYKKGGDASYEPTGPLYLVNVTG-----VDS---KGNELSPH---YVE 46
 Db 347 SVEIDENETKKGGDDYN-----Gp---INVGNLVAFLKIDAEFKSLQCIDPHREYVE 398
 QY 47 -FPIKPGTTLTKKIEYVEWALDATAYKEERVELDPSAKIEVTVYDKNKKEETKSEF- 104
 Db 399 RLDEPQFVLAQNVQYLE-----SIGDFKASSKVALKRVELITYKPHVEYEAIRKLA 452
 QY 105 PTERKGFVVDLSEHKNP-GFN-----LITKVIEKK 136
 Db 453 EMVTEG-----DNGEMSEPEKGFEDTRIPAPFVTVLELVARK 489

RESULT 24
 SYI_CHLUPN STANDARD; PRT; 1043 AA.
 ID SYI_CHLUPN
 AC 09Z972; 09JOK2; 09K221;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Isolation of a tRNA synthetase (EC 6.1.1.5) (Isolation of tRNA ligase)
 DE (IERS).
 GN ILES OR CPN0109 OR CP065.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiophila.
 OX NCBI_TaxID=83536;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CWL029;
 RX MEDLINE-99206506; PubMed-10192388;
 RA Olinger L., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39;
 RX MEDLINE-20150255; PubMed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uettermann T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 RT pneumoniae AR39.";
 RT Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-J138;
 RX MEDLINE-20330349; PubMed-10871362;
 RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + L-Isoleucine + tRNA(Ile) = AMP +
 CC diisophosphate + L-Isoleucyl-tRNA(Ile).
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE001597; AAD18262.1; -
DR EMBL: AE002224; AAF38477.1; -
DR EMBL: AP002545; BAA98320.1; -
DR HSSP: P56690.1; ILE.
DR TIGR: CP0655; -
DR InterPro: IPR002300; tRNA-synt_1a.
DR InterPro: IPR001412; tRNA-synt_1.
DR InterPro: IPR002301; tRNA-synt_1le.
DR Pfam: PF00133; tRNA-synt_1_1.
DR PRINTS: PR00984; TRNASYNTHILE.
DR PROSITE: PS00178; AA-TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; zinc; Complete proteome.
FT SITE 48 58 "HIGH" REGION.
FT SITE 591 595 "KMSKS" REGION.
FT BINDING 594 594 ATP (BY SIMILARITY).
FT CONFLICT 384 384 H -> Y (IN REF. 2).
SQ SEQUENCE 1043 AA; 120286 MW; 08AB8B9364E640C CRC64;

Query Match 10.1%; Score 72.5; DB 1; Length 1043;
Best Local Similarity 20.8%; Pred. No. 53;
Matches 26; Conservative 24; Mismatches 42; Indels 33; Gaps 5;

QY 24 YLAWNVGVDSKGNELSPHYVEPIKPGTLTKETKEIYVE-----W-----AL 68
DB 680 WILSNLYSVGKVESQVHLNFAVEFYTFIDDLNWRKRRKRRFWEDEPDRRAA 739
QY 69 DATYKFRVEELDPASAKIEVY-----YDNKKKEETKS-----FPTKEGFVPP 115
DB 740 FSTLYEVLTVF-----CKVIAPFVPLAEDIYQKLEKEPESVHLCDFQVEMDKILPD 794
QY 116 LSEHI 120
DB 795 LEKRA 799

RESULT 25
YQ30_BACAN STANDARD; PRT; 354 AA.
AC Q9RNO2;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein pXO2-30.
GN pXO2-30.
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
CC NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RA Okinaka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lanke G.,
RA Kunano S., Manter D., Martinez Y., Svensson R., Tatum L.R.,
RA Brown A.E., Jackson P.J.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch).

DR EMBL: AF188935; AAF13635.1; -
KW Hypothetical protein; Plasmid; Transmembrane.
FT TRANSMEM 8 POTENTIAL.
SQ SEQUENCE 354 AA; 39491 MW; F4E5569E00F0F7A CRC64;

Query Match 10.1%; Score 72; DB 1; Length 354;
Best Local Similarity 26.3%; Pred. No. 16;
Matches 36; Conservative 17; Mismatches 50; Indels 26; Gaps 5;

QY 10 KKGDDASYEP-----TGPYLMVNTGVDSKGNELSPHY--VEPIKPGTTLTK 57
DB 69 KKGSDLLDEFAFIKPSDIVDGTPEPDKSNAGNDSSPNNGIYRTFDYTVYPLK--VTIMP 126
QY 58 EKIEIYVENALDTAKKRRVVELDPSAKIEVYIYKNNKKEETKSFPTEKGFVPPDLIS 117
DB 127 KKHKDKLKLKLTGTLENGITNKRVAFAV-----GKED-----LEKGIYSFDM 174
QY 118 EHIKNPGFNLTTRVIE 134
DB 175 YTKKENGNSIMRIVYE 191

RESULT 26
HISX_METUA STANDARD; PRT; 429 AA.
AC O58851;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histidinol dehydrogenase (EC 1.1.1.23) (HDH).
GN HISD OR MJ1456.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
CC NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurl M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: THIS PROTEIN IS CONSIDERED AS A BIFUNCTIONAL ENZYME,
CC POSSESSING TWO ACTIVE SITES, ONE AN ALCOHOL DEHYDROGENASE AND
CC THE OTHER AN ALDEHYDE DEHYDROGENASE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-histidinol + 2 NAD(+) + H(2O) = L-histidine
CC + 2 NADH.
CC -1- PATHWAY: TENTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC, FUNGAL AND PLANTS HDH.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U67586; AAB99465.1; -
DR TIGR: MJ1456; -
DR InterPro: IPR001692; Histidinol_dh.
DR Pfam: PF00815; Histidinol_dh_1.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
 CC PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X61479; CAA43706.1; -
 CC PIR: S16385; S16385.
 CC HSSP: P11362; 1FGK.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR001824; Receptor_Tyr_kin_III.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00047; Ig_4.
 DR Pfam: PF00069; pkinase; 2.
 DR SMART: SM00410; Ig_Like; 3.
 DR SMART: SM00408; IgC2; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00240; RECEPTOR_TYR_KIN_III; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR Proto-oncogene: Tyrosine-protein kinase; Receptor; Transmembrane;
 KW Transferrase; Glycoprotein; Phosphorylation; ATP-binding; Signal;
 KW Immunoglobulin domain; Repeat.
 FT CHAIN 1 19
 FT CHAIN 978
 FT
 FT DOMAIN 20 511
 FT TRANSMEM 512 536
 FT DOMAIN 537 978
 FT DOMAIN 24 104
 FT DOMAIN 107 197
 FT DOMAIN 204 298
 FT DOMAIN 299 397
 FT DOMAIN 398 503
 FT DOMAIN 580 914
 FT NF_BIND 586 594
 FT BINDING 614 614
 FT ACT_SITE 776 776
 FT DISULFID 42 84
 FT DISULFID 127 177
 FT DISULFID 224 278
 FT DISULFID 417 483
 FT MOD_RES 697 697
 FT MOD_RES 706 706
 FT MOD_RES 807 807
 FT MOD_RES
 FT CARBOHYD 45 45
 FT CARBOHYD 73 73
 FT CARBOHYD 302 302
 FT CARBOHYD 335 335
 FT CARBOHYD 389 389
 FT CARBOHYD 410 410
 FT CARBOHYD 449 449
 FT CARBOHYD 478 478
 FT CARBOHYD 491 491
 FT SEQUENCE 978 AA; 109264 MW; 0A68456EF56BC7E3 CRC64;

Query Match 10.18; Score 72; DB 1; Length 978;
 Best Local Similarity 23.28; Pred. No. 54;

Matches 35; Conservative 20; Mismatches 44; Indels 52; Gaps 6;
 QY 12 GDDASYEPTGYLWVN-----VTGVDSKGNELSPHYVEPIKPTTLKREK--- 59
 Db 17 GCGAVIEPSPGELVEEGEYTLACVNSGVEMDGPISPTWTLDPESPSTLTTRNATF 76
 QY 60 -----IEYVE-----WALDATAKFERVE-----LDPS 84
 Db 77 KWTGTYRCELEDPMAAGSTTHLYKDPANSHNLLA---GEYTVVEGEAVLPCLITDPA 133
 QY 85 AKIEYTYIDKNNKK--EEKSPITEKGFV 113
 Db 134 LKDSVSLNKEGGROVLRRTYFFESAMRGFTI 164
 RESULT 29
 ID YE20_METJA STANDARD; PRT; 1102 AA.
 AC 058815;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein M01420 [contains: Mja gfpG Intein].
 GN M01420.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 ON NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Georgagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Weese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -1- P1M: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
 CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
 CC -1- SIMILARITY: IN THE C-TERMINAL, NO GLUCOSAMINE--FRUCTOSE-6-
 CC PHOSPHATE AMINOTRANSFERASES (GFAT).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U67582; AAB99430.1; -
 CC HSSP: P17169; 1MOO.
 DR MEROPS: C44.971; -
 DR TIGR: M01420; -
 DR InterPro: IPR000583; GATase_2.
 DR InterPro: IPR001387; HTH_3.
 DR InterPro: IPR003586; HntC.
 DR InterPro: IPR003587; HntN.
 DR InterPro: IPR002203; Intein.
 DR InterPro: IPR004042; Intein_endonuc.
 DR InterPro: IPR001347; SIS.
 DR Pfam: PF00310; GATase_2; 2.
 DR Pfam: PF01381; HTH_3; 1.
 DR Pfam: PF01380; SIS; 2.
 DR SMART: SM00530; HTH_XRE; 1.
 DR SMART: SM00305; HntC; 1.

DR SMART; SM00306; H1TEN; 1.
 DR PROSITE; PS00443; GANASE_TYPE_II; 1.
 DR PROSITE; PS00881; PROTEIN_SPLICING; FALSE_NEG.
 KM Hypothetical protein; Autocatalytic cleavage; Protein splicing;
 KM Complete proteome.
 FT CHAIN 1 74 HYPOTHETICAL PROTEIN MJ1420, 1ST PART
 FT CHAIN 75 573 (POTENTIAL).
 FT CHAIN 574 1102 (POTENTIAL).
 FT CHAIN 574 1102 HYPOTHETICAL PROTEIN MJ1420, 2ND PART
 FT CHAIN 574 1102 (POTENTIAL).
 SQ SEQUENCE 1102 AA; 125907 MW; 9C1EFA88D90CB1 CRC64;

Query Match 10.1%; Score 72; DB 1; Length 1102;
 Best Local Similarity 25.5%; Pred. No. 62;
 Matches 41; Conservative 21; Mismatches 51; Indels 48; Gaps 11;

QY 6 KCKYKGGDASY-FPEPTGYLM---VNTGYDSK-GNELSPHYVEPIKGTLTKEK 59
 DB 321 EGNKKGDENYIILEINSKYLIDMERENIPELFNKTGNE-KTPEFY-----FRLNDL 372
 QY 60 IEYVEMALDATAY--KEFRVELDPSAKI---EVTY-----YDKNRKEET--- 101
 DB 373 VASYLNGIFDAGCYIRAEAKQIGWTSKCFKEIQFLRLRGILASYSKIKRKEENNN 432
 QY 102 -----KSPITEK-GEVVPDLSEHI-----KNGCFN 126
 DB 433 THKLISDKSELEFKYIGFTAKDKMERLEALNKMGLN 473

RESULT 30
 ID CENE_HUMAN STANDARD; PRT; 2663 AA.
 AC Q02224;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Centromeric protein E (CENP-E protein).
 GN CENPE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=106971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Thirmer D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RT microtubule motor";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and hBUBR1";
 RL Cell Biol. 143:49-63(1998).
 CC -1- FUNCTION: KINOS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z15005; CAAT8727.1; .
 DR PIR; S28261; S28261.
 DR HSSP; P17119; 3KAR.
 DR MTM; 117143; .
 DR InterPro; IPR001752; Kinesin.
 DR Pfam; PF00225; Kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN; 1.
 DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN; 1.
 KM Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 KW Cell cycle; Centromere.
 FT DOMAIN 1 335 KINESIN-MOTOR.
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT NP_BIND 86 93 ATP (BY SIMILARITY).
 SQ SEQUENCE 2663 AA; 312087 MW; CFCJ3880C8C68 CRC64;

Query Match 10.1%; Score 72; DB 1; Length 2663;
 Best Local Similarity 28.7%; Pred. No. 18+02;
 Matches 25; Conservative 16; Mismatches 26; Indels 20; Gaps 4;

QY 27 VNTGYDSKGNELSPHYVEPIKGTLTKEK---EYVEMALDATAYKEFRVEL-- 81
 DB 2321 VNPYTDNKK-----NPHVT-----SRATQLTTEKIRELENSLHKEAKESAMKREKIIKMK 2371
 QY 82 -----DPSAKIEVTYDKNRKEETK 102
 DB 2372 ELEVYNDIATLAKVHESNCKLEKTK 2398

RESULT 31
 ID YF6L_METJA STANDARD; PRT; 553 AA.
 AC Q38936;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ1561.
 GN MJ1561.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake O., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii";
 RL Science 273:1056-1073(1996).
 CC -1- SIMILARITY: TO M.JANNA SCHII MJ0795 AND MJ1506.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U67596; AAB99582.1; -.
 DR TIGR; M1561; -.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 6 26 POTENTIAL.
 FT TRANSMEM 524 544 POTENTIAL.
 SQ SEQUENCE 553 AA; 61713 MW; 86CA1E67C5A0AA01 CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 553;
 Best Local Similarity 23.1%; Pred. No. 31;
 Matches 36; Conservative 24; Mismatches 43; Indels 53; Gaps 9;

QY 7 GKTKGD-----DASYFEPTGYLMVNTGVDSKG-----NELLSPH 43
 DB 283 GAYKVGDIILGISNVYDPEIKRGTYVRIDVT-ITNGHAEAKDVKLITNKPKFDS 341
 QY 44 YVERPIK-----PGTTLTKEKIEYVEMALDATAYKEFVVELDPSAKIEVYYDKNK 97
 DB 342 WSNONIDVGNLPGVSKT---VSFYVDVKYASA-KHYKL-----PLEISYLDITANN 390
 QY 98 KEETKSPITEKGFVVDLSEHIK-NPGFNILITKVY 132
 DB 391 KYTKERF-----IDIVYKPKLEFIIITKEV 415

RESULT 32
 ID POTL_MYCPN STANDARD; PRT; 560 AA.
 AC P75059;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Spermidine/putrescine transport ATP-binding protein potA homolog.
 OS POTA OR MPN055 OR MP099.
 GN Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE-97105885; PubMed-8948633;
 RA Himmelfeich R., Hilbert H., Plagens H., Pirk E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC FOR SPERMIDINE/PUTRESCINE. PROBABLY RESPONSIBLE FOR ENERGY
 CC COUPLING TO THE TRANSPORT SYSTEM (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AE000011; AAB95747.1; -.
 DR HSSP; P13569; INBD.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR003439; ABC_transportr.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam; PF00005; ABC_tran; 1.

DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW Transport; ATP-binding; Membrane; Complete proteome.
 FT NP_BIND 40 47 ATP (POTENTIAL).
 SQ SEQUENCE 560 AA; 65131 MW; 8993CAE61A28723D CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 560;
 Best Local Similarity 22.5%; Pred. No. 31;
 Matches 39; Conservative 22; Mismatches 51; Indels 61; Gaps 9;

QY 1 SSSFDKG-----KTKGDASYFEPTG-----PYLMVNTGVDSK 35
 DB 13 NRTFDDGFVSVRDINLKIKKGEFIIILGPGCGKTYTLLLAGFEDPYRGKIKVGLDIK 72
 QY 36 GNEILSPHVERPI--KPGTTLTKEKIEYVEMAL--DATAKER-----RVVELDPSA 85
 DB 73 -----DLPIHKRPATV-----FDYALFSLTVYKNIAVGIKAMYTKLDPID 115
 QY 86 KIEVYYDK--NKKKEETKSPITEKGFVVDLSEHIK-NPGFNILITKVIEKK 136
 DB 116 KLVEQYHSLDKQHRHLRIERLEKSNANPQLDLQAK-----TYVQOK 161

RESULT 33
 ID Y366_MYCGE STANDARD; PRT; 667 AA.
 AC P47606;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MG366.
 GN MG366.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE-96026346; PubMed-7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrman J.L.,
 RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U39718; AAC71593.1; -.
 DR TIGR; MG366; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 667 AA; 78117 MW; 450389698D0690D8 CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 667;
 Best Local Similarity 26.5%; Pred. No. 38;
 Matches 27; Conservative 16; Mismatches 40; Indels 19; Gaps 4;

QY 46 EPIKPGTTLTKEKIEYVEMALDATAYKE-----FVVELDPSAKIEVYYDKN 95
 DB 162 EFVYKATITNTRK--YFEIIYDFLFEKKIKKYLANTFCTVAYEIQNKNNVSFFLNT 219
 QY 96 KKEETKSPITEKGFVVDLSEHIK-NPGFNILITKVY-IEKK 136

Db 220 EIKTSKNSFSLSSKE-----KDYFNKPFNHEKAIYTHKK 255

RESULT 34

FREL_YEAST 34

ID FREL_YEAST STANDARD: PRT: 686 AA.

AC P32791.

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ferric reductase transmembrane component 1 precursor (EC 1.6.99.13)

GN (Ferric-chelate reductase 1).

GN FREL OR YLR214W OR L8167.2.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=F113;

RC MEDLINE=92237270; PubMed=1570306;

RA Dancis A., Roman D.G., Anderson G.J., Hanebusch A.G., Klausner R.D.;

RT "Ferric reductase of *Saccharomyces cerevisiae*: molecular

RT characterization, role in iron uptake, and transcriptional control by

RT iron."

RL Proc. Natl. Acad. Sci. U.S.A. 89:3869-3873(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,

RA Favallo A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,

RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,

RA Johnston L., Langston Y., Latreille P., Mardis E., Meneses S.,

RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,

RA Tatch S., Trevaaskis E., Vaudin M., Vignati D., Wilcox L., Wilson R.,

RA Woldman P., Waterston R.;

RL Submitted (SEP-1994) to the EMBL/GenBank/DDJB databases.

RN [3]

RP CHARACTERIZATION.

RC MEDLINE=93057491; PubMed=1431884;

RA Anderson G.J., Lesuisse E., Dancis A., Roman D.G., Labbe P.,

RA Klausner R.D.;

RT "Ferric iron reduction and iron assimilation in *Saccharomyces*

RT *cerevisiae*."

RL J. Inorg. Biochem. 47:249-255(1992).

CC -1- FUNCTION: REDUCTASE ACTIVITY THAT ACTS ON FERRIC IRON CHELATES

CC EXTERNAL TO THE CELL. PLAYS A ROLE IN IRON UPTAKE. MAY

CC PARTICIPATE IN THE TRANSPORT OF ELECTRONS FROM CYTOPLASM TO AN

CC EXTRACELLULAR SUBSTRATE (FERRIC ION) VIA FAD AND HEME

CC INTERMEDIATES. MAY ALSO PARTICIPATE IN CU(II) REDUCTION AND CU(I)

CC UPTAKE.

CC -1- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).

CC -1- COFACTOR: FAD (PROBABLE).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- INDUCTION: BY IRON DEPRIVATION. REPRRESSD BY IRON UPTAKE.

CC -1- SIMILARITY: BELONGS TO THE FREL / CYB FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: M86908; AAA34608.1; -

CC EMBL: U14913; AAB67424.1; -

CC PIR: S30075; S30075.

CC PIR: S48565; S48565.

CC SGD: S0004204; FREL.

CC InterPro: IPR002916; Ferric_reduct.

CC Pfam: PF01794; Ferric_reduct; 1.

KW Oxidoreductase: Electron transport; Transmembrane; Iron transport;

KW FAD; NAD; Copper; Glycoprotein; Signal; Multigene family.

FT SIGNAL 1

FT CHAIN 23

FT NP_BIND 462 468 FAD (POTENTIAL).

FT NP_BIND 532 540 NAD (POTENTIAL).

FT TRANSMEM 147 169 1 (POTENTIAL).

FT TRANSMEM 216 236 2 (POTENTIAL).

FT TRANSMEM 258 277 3 (POTENTIAL).

FT TRANSMEM 296 316 4 (POTENTIAL).

FT TRANSMEM 329 348 5 (POTENTIAL).

FT TRANSMEM 369 397 6 (POTENTIAL).

FT TRANSMEM 329 350 7 (POTENTIAL).

FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 100 104 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 686 AA; 78853 MW; 7F6BB3B93A95D6A3 CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 686;

Best Local Similarity 23.0%; Pred. No. 40;

Matches 29; Conservative 24; Mismatches 44; Indels 29; Gaps 6;

OY 22 GPY-----LMVNTGVS-KGNELSPHYEPFPIKGTTLTKREKIEVYEMLDAT 71

DB 514 GPYGVTVPHIAKRLKRLVGAAGLVAALYHPHEVCLRLPST-----DQJHKRYWVNDL 569

OY 72 AYKEFRVVEL-----DPSAKIEVTY-----YDKNKKKEETSFPTTEGFFV-----PDL 116

DB 570 SHKMFENELQWLKESCVSYITGSSVEDTNSDESTGFDKDESEITVECLNRPDL 629

OY 117 SEHKN 122

DB 630 KELVRS 635

RESULT 35

CYSA_SYNP7

ID CYSA_SYNP7 STANDARD: PRT: 344 AA.

AC P14788;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sulfate transport ATP-binding protein *cybA*.

GN CYSA.

OS *Synechococcus* sp. (strain PCC 7942) (*Anaerostis nidulans* R2).

OC Bacteria; Cyanobacteria; Chroococcales; *Synechococcus*.

OX NCBI_TaxID=1140;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=89184544; PubMed=2538823;

RA Green L.S., Laudenbach D.E., Grossman A.R.;

RT "A region of a cyanobacterial genome required for sulfate transport."

RL Proc. Natl. Acad. Sci. U.S.A. 86:1949-1953(1989).

RN [2]

RP SEQUENCE OF 1-11 FROM N.A.

RC MEDLINE=91210162; PubMed=1708375;

RA Laudenbach D.E., Grossman A.R.;

RT "Characterization and mutagenesis of sulfur-regulated genes in a

RT cyanobacterium: evidence for function in sulfate transport."

RL J. Bacteriol. 173:2739-2750(1991).

CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF MULTIPLE SULFUR-CONTAINING

CC COMPOUNDS, INCLUDING SULFATE AND THIOSULFATE.

CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.

CC -1- INDUCTION: BY SULFUR DEPRIVATION.

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb.ch/announce>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; J04512; AAA22056.1; -.
DR EMBL; M65247; AAA73042.2; -.
DR PIR; A30301; GRTCS7.
DR PIR; J00123; J00123.
DR HSSP; P13569; INBD.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Sulfate transport; Transport; ATP-binding; Inner membrane.
FT NP_BIND 41 48 ATP (BY SIMILARITY).
SQ SEQUENCE 344 AA; 38476 MW; 1A7781AD0A8A55 CMC64;

```

CC -1- MISCELLANEOUS: 'Croquemort' means literally the one who bites
CC dead persons', in French and is colloquial for undertaker.
CC -1- SIMILARITY: BELONGS TO THE CD36 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC -----
DR DR EMBL; Z31583; CAA83455.1; -.
DR DR EMBL; Z31582; CAA83454.1; -.
DR DR FLYBase; FBgn0015924; crg.
DR DR InterPro; IPR002159; CD36.
DR DR Pfam; PF01130; CD36; 1.
KM Glycoprotein; Transmembrane; Apoptosis; Palmitate; Lipoprotein;
KW Receptor.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).

```

Query Match	9.9%	Score 71;	DB 1;	Length 344;
Best Local Similarity	23.1%	Pred. No. 19;		
Matches	28;	Conservative	20;	Mismatches 35; Indels 38; Gaps 6;
QY	17	YEEPPGYPLM-----VNVT-----	GVDSKGNEL-TSPHYVEFPKPGTTLTKERI	60
Db	227	YDNPATPFVMSITGIVNVLPNSHITFGAGIDTTPRPYFLNRHDEIMADLPETVPARI		286
QY	61	EYVE--WALDATATKE-----PRVELEDSAKIEVTYYDKKKKEETKSEF		105
Db	287	DRIVHIGVEVQAEVRLTEDGQVLAHLPRDPRYEDLTLEEQGVFV-----RPKQARSEF		339
QY	106	I 106		
Db	340	L 340		

FT	LIPID	2	2	PALMITATE (POTENTIAL).
FT	LIPID	3	3	PALMITATE (POTENTIAL).
FT	LIPID	5	5	PALMITATE (POTENTIAL).
FT	LIPID	6	6	PALMITATE (POTENTIAL).
FT	DISULFID	256	324	BY SIMILARITY.
FT	DISULFID	285	345	BY SIMILARITY.
FT	DISULFID	326	334	BY SIMILARITY.
FT	CARBOHYD	72	72	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	104	104	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	110	110	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	129	129	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	189	189	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	198	198	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	216	216	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	260	260	N-LINKED (GLCNAC. . .)
SO	SEQUENCE	457 AA;	52318 MW;	602E86AC1EF72736D CRC64;

```

RESULT 36
CRO_QDROME: STANDARD: PRT: 457 AA.
ID CRO_QDROME:
AC 02/36/7;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Croquemort protein (D-CD36).
CRO:
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
XX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RA MEDLINE=96222807; PubMed=8630729;
RX Franc N.C., Dimaecq J.-L., Lagueny M., Hoffmann J., Ezekowitz R.A.B.;
RT "Croquemort, a novel Drosophila hemocyte/macrophage receptor that
RT recognizes apoptotic cells.";
RL Immunity 4:431-443(1996).
CC -1- FUNCTION: MACROPHAGE RECEPTOR FOR APOPTOTIC CELLS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: FOUND ON HEMOCLYMPH, AND LATER
CC IN THE DEVELOPMENT, ON MACROPHAGES THAT CONTAIN APOPTOTIC CELL
CC CORPSES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ON HEMOCYTES/MACROPHAGES IN
CC EMBRYOGENESIS. FIRST DETECTED IN LATE STAGE 11 EMBRYOS, WHICH
CC CORRESPONDS TO THE FIRST WAVE OF APOPTOSIS. EXPRESSED IN CELLS
CC THAT BY STAGE 12 SPREAD THROUGHOUT THE EMBRYO AND BY STAGE 13/14
CC HAVE MIGRATED FROM BOTH ENDS TOWARDS THE MIDDLE UNTIL THEY ARE
CC EVENLY DISTRIBUTED BY STAGE 15.
CC -1- PTM: N-GLYCOSYLATED AND/OR O-GLYCOSYLATED (PROBABLE).

```

```

Query Match          9.9%; Score 71; DB 1; Length 457;
Best Local Similarity 25.5%; Pred. No. 27;
Matches 35; Conservative 18; Mismatches 56; Indels 28; Gaps 5;

QY      10 KKGDAA--SYEPENGR-----YLMVNVGVDSKGNELSPHYVEPIRKQDTLTREKIEY 63
          | | - - - - - | | | | | | | | | | | | | | | | | | | | | |
DB      48 KPGDDAYSSMLBAPLPYLSFYEMNWTNPEDIRNDIPNFE--MGPYTFLEKHKENY 105
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      64 VEMALDAAAYAKKEFRVELDP-----SAKIEVYYDDNNKKKEEKSEPI 106
          : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      106 TFYDNATVAAVEYRRTFWTFDPERSNCTLDWVTAAHATATVADEKRDQRIKIKKLIINFL 165
          : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      107 TEKG---FVDPDLSEHI 120
          : | : | : | : | |
DB      166 NHEGGLVYTRKPVGEWI 182

RESULT 37
RFBM_SALTY
ID      RFBM_SALTY      STANDARD:      PRT:      479 AA.
AC      P26404;
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Mannose-1-phosphate guanylyltransferase (GDP) (EC 2.7.7.22) (GDP-
      mannose-1-phosphorylase) (GMP).
RFBM OR STM2084.
OS      Salmonella typhimurium.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Salmonella.
OX      NCBI_TaxID=602;
RN      [1]
RP      SEQUENCE FROM N.A.

```

CC STRAIN-LT2;
 RA MEDLINE-91260454; PubMed-1710759;
 RA Jiang X.-M., Neal B., Santiago F., Lee S.J., Romana L.K., Reeves P.R.;
 RT "Structure and sequence of the rfb (O antigen) gene cluster of
 RT Salmonella serovar typhimurium (strain LT2).";
 RL Mol. Microbiol. 5:695-713(1991).
 RM [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE-21534948; PubMed-11677609;
 RA McClelland M., Sanderson K.E., Speth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Miliyanev E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: INVOLVED IN GDP-MANNOSE BIOSYNTHESIS WHICH SERVES AS
 CC THE ACTIVATED SUGAR NUCLEOTIDE PRECURSOR FOR MANNOSE RESIDUES
 CC IN CELL SURFACE POLYSACCHARIDES. THIS ENZYME PARTICIPATES IN
 CC SYNTHESIS OF THE LPS GROUP B O ANTIGEN.
 CC -1- CATALYTIC ACTIVITY: GDP + D-mannose 1-phosphate = phosphate + GDP-
 CC mannose.
 CC -1- PATHWAY: GDP-MANNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
 CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF MANNOSE-6-PHOSPHATE ISOMERASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X56793; GAA40128.1; -
 CC EMBL: AF008792; AAL20986.1; -
 CC PIR: S15312; S15312.
 CC StyGene; SG10350; rfbM.
 CC InterPro: IPR001538; MannoseP_Isomer.
 CC InterPro: IPR001825; NTP_transferase.
 CC Pfam: PF01050; MannoseP_Isomer; 1.
 CC Pfam: PF00483; NTP_transferase; 1.
 CC ProDom: PD002664; MannoseP_Isomer; 1.
 CC DR Lipopolysaccharide biosynthesis; Transferase; Kinase;
 CC KW Complete proteome.
 CC SQ SEQUENCE 479 AA; 54046 MW; FE9E84D5CEEF118 CRC64;

Query Match 9.9%; Score 71; DB 1; Length 479;
 Best Local Similarity 22.9%; Pred. No. 29;
 Matches 30; Conservative 21; Mismatches 48; Indels 32; Gaps 7;

OY 5 DKGKYGKGDASYFEPTGYLMVNTGVDSKGNELSPHYVEPIKPGTTLKKEIEYV 64
 DB 352 DKEKYGQDEV--FRPWGTY--NW--IDSGKNYLVRC---ITYKPGKEFYAQNHNHNA 400
 OY 65 E-WALDATTAKKFEFVVELDPSAKIEVYDYDNKKKEETKSPFTEKGFVVDLSEHKNP 123
 DB 401 EHM-----IVLSGTARV-----TKGEQTFWVSENSSTLIPPTIALENP 440
 OY 124 GENLITKVVIE 134
 DB 441 GMTPLKLEIQ 451

RESULT 38
 PUR9_BACSU STANDARD: PRT; 512 AA.
 AC P12048;
 DT 01-OCT-1989 (rel. 12, Created)
 DT 01-OCT-1989 (rel. 12, Last sequence update)

DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Bifunctional purine biosynthesis protein purH [includes:
 DE Phosphoribosylamidoimidazolecarboxamide formyltransferase (EC 2.1.2.3)
 DE (AICAR transformylase); IMP cyclohydroxylase (EC 3.5.4.10) (Inosinase)
 DE (IMP synthetase) (ATC)].
 GN PURH OR PURH2.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RP SEQUENCE FROM N.A.
 RX MEDLINE-67250425; PubMed-3036807;
 RA Ebdole D.J., Zalkin H.;
 RT "Cloning and characterization of a 12-gene cluster from Bacillus
 RT subtilis encoding nine enzymes for de novo purine nucleotide
 RT synthesis.";
 RL J. Biol. Chem. 262:8274-8287(1987).
 RN [2]
 RP SEQUENCE OF 432-512 FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-97124186; PubMed-8969499;
 RA Borries R., Porwollik S., Schroeter R.;
 RT "The 52 degrees-55 degrees segment of the Bacillus subtilis
 RT chromosome: a region devoted to purine uptake and metabolism, and
 RT containing the genes cotA, gabp and guaA and the pur gene cluster
 RT within a 34960 bp nucleotide sequence.";
 RL Microbiology 142:3027-3031(1996).
 CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5'-
 CC phosphoribosyl-5-amino-4-imidazolecarboxamide = tetrahydrofolate +
 CC 5-phosphoribosyl-5-formamido-4-imidazolecarboxamide.
 CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-
 CC phosphoribosyl)imidazole-4-carboxamide.
 CC -1- PATHWAY: NINTH AND TENTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
 CC -1- DOMAIN: THE IMP CYCLOHYDROLASE ACTIVITY RESIDES IN THE N-TERMINAL
 CC REGION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PURH FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J02732; AAA22683.1; -
 CC EMBL: AF011544; AAB72185.1; -
 CC EMBL: Z99107; CAB12472.1; -
 CC PIR: A29183; DTRSPH.
 CC DR Subtilist; BG10710; purH.
 CC InterPro: IPR002695; AICARFT_IMPChas.
 CC InterPro: IPR000384; MGS.
 CC Pfam: PF01808; AICARFT_IMPChas; 1.
 CC Pfam: PF02142; MGS; 1.
 CC ProDom: PD004666; AICARFT_IMPChas; 1.
 CC DR Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme;
 CC KW Complete proteome.
 CC SQ SEQUENCE 512 AA; 55739 MW; 7FD6B0FE5485234C CRC64;

Query Match 9.9%; Score 71; DB 1; Length 512;
 Best Local Similarity 27.6%; Pred. No. 31;
 Matches 35; Conservative 19; Mismatches 41; Indels 32; Gaps 8;

OY 1 SSSFDKGYKKGDASYFEPTGYLMVNTGVDSKGNELSPHYVEPIKPGTTLKKEKI 60
 DB 288 AEAFDRA--FEADKTSIF--GGITALN-REVDATAEALNIFLETITAPS----- 333
 OY 61 EYVEMALDA-TAKKFEFVVELDPSAKIEVYDYDNKKKEETKSPFTEKGFVVDLSEH 119
 DB 334 --FQGEALDVTAKKNRLIVTLVDYSAVQ-----KEKQUTS-----VQGGLLIQDLDMH 380

QY 120 IKMPGN 126
 DB 381 ----GFD 383
 RESULT 39
 KFM_MOUSE STANDARD; PRT; 976 AA.
 ID KFM_MOUSE
 AC P09581;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DE Macrophage colony stimulating factor 1 receptor precursor (CSF-1-R)
 DE (EC 2.7.1.112) (fms proto-oncogene) (c-fms).
 OS CSF1R OR CSFMR OR FMS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX MEDLINE-88217329; PubMed-2966922;
 RA Rothwell V.M., Rohrschneider L.R.;
 RT "Murine c-fms cDNA: cloning, sequence analysis and retroviral
 expression.";
 RL Oncogene Res. 1:311-324(1987).
 RN (2)
 RP REVIEWS.
 RA Rothwell V.M.;
 RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP REVIEWS.
 RA MEDLINE-93181280; PubMed-8441691;
 RA de Parseval N., Borderaux D., Gisselbrecht S., Sola B.;
 RT "Assessment of the murine c-fms proto-oncogene sequence.";
 RL Nucleic Acids Res. 21:750-750(1993).
 RN (4)
 RP SEQUENCE OF 1-16 FROM N.A.
 RX MEDLINE-93268269; PubMed-8497248;
 RA Yue X., Favot P., Dunn T.L., Cassady A.I., Hume D.A.;
 RT "Expression of mRNA encoding the macrophage colony-stimulating factor
 receptor (c-fms) is controlled by a constitutive promoter and tissue-
 specific transcription elongation.";
 RL Mol. Cell. Biol. 13:3191-3201(1993).
 RN (5)
 RP AUTOPHOSPHORYLATION SITES.
 RX MEDLINE-90258890; PubMed-2160591;
 RA van der Geer P., Hunter T.;
 RT "Identification of tyrosine 706 in the kinase insert as the major
 colony-stimulating factor 1 (CSF-1)-stimulated autophosphorylation
 site in the CSF-1 receptor in a murine macrophage cell line.";
 RL Mol. Cell. Biol. 10:2991-3002(1990).
 CC -1- FUNCTION: THIS PROTEIN IS THE RECEPTOR FOR CSF-1, IT IS A PROTEIN
 CC TYROSINE-KINASE TRANSMEMBRANE RECEPTOR.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine -> ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE CSF-1/PDGFR RECEPTOR FAMILY OF TYROSINE-
 CC PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X06368; CAA29666.1; ALT_SEQ.
 CC EMBL; S62219; -; NOT_ANNOTATED_CDS.
 CC PIR; S01880; TYMSMD.
 CC HSSP; P11362; IFGK.

DR MGD: MGI:1339758; Csflr.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR001824; Receptor_tyr_kin_III.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00047; Ig_4.
 DR Pfam: PF00069; pkinase; 2.
 DR SMART: SM00410; IG_1like; 3.
 DR SMART: SM00408; IGC2; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00240; RECEPTOR_TYR_KIN_III; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Proto-oncogene: Tyrosine-protein kinase; Receptor; Transmembrane;
 KW Transferase; Glycoprotein; Phosphorylation; ATP-binding; Signal;
 KW Immunoglobulin domain; Repeat.
 FT CHAIN 1 19
 FT SIGNAL 20 976
 FT
 FT DOMAIN 20 511
 FT TRANSMEM 512 536
 FT DOMAIN 537 976
 FT DOMAIN 24 104
 FT DOMAIN 107 197
 FT DOMAIN 204 298
 FT DOMAIN 289 397
 FT DOMAIN 398 503
 FT DOMAIN 580 913
 FT NP_BIND 586 594
 FT BINDING 614 614
 FT ACT_SITE 776 776
 FT DISULFID 42 84
 FT DISULFID 127 177
 FT DISULFID 224 278
 FT DISULFID 417 483
 FT MOD_RES 697 697
 FT MOD_RES 706 706
 FT MOD_RES 807 807
 FT CARBOHYD 45 45
 FT CARBOHYD 73 73
 FT CARBOHYD 302 302
 FT CARBOHYD 335 335
 FT CARBOHYD 389 389
 FT CARBOHYD 410 410
 FT CARBOHYD 449 449
 FT CARBOHYD 478 478
 FT CARBOHYD 491 491
 SQ SEQUENCE 976 AA; 109020 MM; AC03051037F6A28D CRC64;
 Query Match 9.9%; Score 71; DB 1; Length 976;
 Best Local Similarity 23.2%; Pred. No. 67;
 Matches 35; Conservative 20; Mismatches 44; Indels 52; Gaps 6;
 QY 12 GDASFEETGPTLWYN-----VTGVDSKGNELSPHVEPIKGTLYTEK--- 59
 DB 17 GCGAPVIEFSGELVVEPEYTLKCVSNGSVEMDQIPSYTLDPESGSLITRNATF 76
 QY 60 -----IEYVE-----MALDATAKEEYVE-----LDPS 84
 DB 77 KNTGYRCTELEDPMGSTTHLYVDPAHNNLLA---QEVTVGGEAVLPCLITDPA 133
 QY 85 AKLEVITYDKNKKK--EETKSPITEKGVV 113
 DB 134 LKDSVSLMREGROVLKRTVTFEFSFWRGPII 164
 RESULT 40
 YPT7_CAEEL STANDARD; PRT; 1021 AA.
 ID YPT7_CAEEL

```

AC P41885;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 111.7 kDa protein P37A4.7 in chromosome III.
GN P37A4.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ.
RA Fulton L., Waterston R.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.
CC -1- SIMILARITY: SOME, TO YEAST YNL087M.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00032; AAA50635.1; -
CC HSSP; P47709; 12BD.
CC WormPep; F37A4.7; CE00705.
CC InterPro; IPR000008; C2.
CC InterPro; IPR002149; IRT.
CC InterPro; IPR003315; RPH3A_effector.
CC InterPro; IPR001565; Synaptolegmin.
CC InterPro; IPR000306; Znf_FYVE.
CC Pfam; PF00168; C2; 2.
CC Pfam; PF02318; RPH3A_effector; 1.
CC PRINTS; PR00360; C2DOMAIN.
CC PRINTS; PR00399; SYNAPTOLEGMIN.
CC SMART; SM00239; C2; 2.
CC PROSITE; PS00499; C2_DOMAIN_1; 2.
CC PROSITE; PS50004; C2_DOMAIN_2; 2.
CC PROSITE; PS50178; ZF_FYVE; 1.
CC Hypothetical protein: Repeat; Zinc-finger.
CC ZN_FING 168 233 FIVE-TYPE.
CC DOMAIN 758 847 C2 DOMAIN 1.
CC DOMAIN 898 987 C2 DOMAIN 2.
CC SEQUENCE 1021 AA; 111749 MW; 947C38B7C37BA462 CRC64;

Query Match 9.9%; Score 71; DB 1; Length 1021;
Best Local Similarity 28.6%; Pred. NO. 71;
Matches 34; Conservative 16; Mismatches 39; Indels 30; Gaps 8;

OY 28 NYTVGDSKGNELSPHYVEFPKPGT---LTKKIEYV--EWALATAYKRRVEL 81
DB 768 NIKAMDSC---ESDPYKAFHLLPNTKATKLTSTIKTLNPEKNEMSY---GLED 821
OY 82 DSAKI-EVTVYDNK-----KK---EETKSPFI-TEKGFVVDLSEHIN 122
DB 822 DKKILRTVTVDRDRIGSDFLGRTIALKKLNDENMKFNLYESALPVQGTKEEN 880

RESULT 41
TCPG_YEAST STANDARD; PRT; 534 AA.
ID TCPG_YEAST
AC P39077;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE T-complex protein 1, gamma subunit (TCP-1-gamma) (CCP-gamma).
DE CC13 OR TCP3 OR BIN2 OR YJL014W OR J1336.
DE Saccharomyces cerevisiae (Baker's yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94377502; PubMed=7916460;
RA Chen X., Sullivan D.S., Huftaker T.C.;
RT "Two yeast genes with similarity to TCP-1 are required for
RT microtubule and actin function in vivo."
RT Proc. Natl. Acad. Sci. U.S.A. 91:9111-9115(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA To Van D., Perea J., Jacq C.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF
CC ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
CC ACTIN AND TUBULIN. IN YEAST MAY PLAY A ROLE IN MITOTIC SPINDLE
CC FORMATION.
CC -1- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 KDA THAT
CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U09480; AAA21658.1; -
CC EMBL; Z45289; CAAB9305.1; -
CC PIR; S48875; S48875.
CC HSSP; P48425; 1A6D.
CC SGD; S0003551; CCT3.
CC InterPro; IPR002423; TCP1_cpn60.
CC InterPro; IPR002194; TCP_1.
CC Pfam; PF00118; cpn60_TCP1.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00750; TCP1_1; 1.
CC PROSITE; PS00751; TCP1_2; 1.
CC PROSITE; PS00995; TCP1_3; 1.
CC Chapterone; ATP-binding; Multigene family.
CC K Chapterone; ATP-binding; Multigene family.
CC FT CONFLICT 25 26 IT -> HA (IN REF. 1).
CC FT CONFLICT 86 89 DEEV -> ERKG (IN REF. 1).
CC FT CONFLICT 111 111 L -> F (IN REF. 1).
CC FT CONFLICT 265 265 K -> M (IN REF. 1).
CC FT CONFLICT 274 274 O -> L (IN REF. 1).
CC FT CONFLICT 292 292 MISSING (IN REF. 1).
CC FT CONFLICT 478 478 G -> A (IN REF. 1).
CC SEQUENCE 534 AA; 58814 MW; 5C5AF3D67D9A3B6 CRC64;

Query Match 9.9%; Score 70.5; DB 1; Length 534;
Best Local Similarity 25.7%; Pred. NO. 36;
Matches 27; Conservative 17; Mismatches 32; Indels 29; Gaps 5;

OY 61 EYVEW-----ALDA--TAYKEF-RVVELDPSAKIEVTVYDNK-----KEETKS 103
DB 162 KYVTHMEKKMCELALDAVKVRKDLGTVGSEPNFELDRIYRVYRKIGGVDLSRYLK 221
OY 104 PPTKEGFVVDLSEHINKPGLI-----TKVIEKK 136
DB 222 GVLNKKDVVHPKMSRIENBRVVLDCPLEYKKGESQTNIEIKK 266

RESULT 42
DNLI_PYRKO STANDARD; PRT; 559 AA.
ID DNLI_PYRKO
AC O9HNC4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```


DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thermosstable DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide synthase
 DE (ATP)) (lig(Tk)).
 GN lig.
 OS Pyrococcus kodakarensis.
 OC Archaea: Euryarchaeota: Thermococcales: Thermococcaceae: Thermococcus.
 ON NCBI_TaxID=69014;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=KOD1.
 RX MEDLINE=20507812; PubMed=11053387;
 RA Nakatani M., Ezaki S., Atomi H., Imanaka T.;
 RA A DNA ligase from a hyperthermophilic archaeon with unique cofactor
 RT specificity.*;
 RL J. Bacteriol. 182:6424-6433(2000).
 CC -1- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA
 CC RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.
 CC -1- CATALYTIC ACTIVITY: ATP + (deoxyribonucleotide)(N) +
 CC (deoxyribonucleotide)(N) -> AMP + diphosphate +
 CC {deoxyribonucleotide}(N-M).
 CC -1- COFACTOR: REQUIRES MAGNESIUM. CAN ALSO ACTS WITH NAD AS A
 CC COFACTOR.
 CC -1- SUBUNIT: MONOMER.
 CC -1- MISCELLANEOUS: OPTIMUM PH IS 8.0. STILL ACTIVE AT 100 DEGREES
 CC CELSIUS.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AB042527; BAB1949.1; ALT.INT.
 DR InterPro: IPR000977; DNA_Ligase.
 DR Pfam: PF01068; DNA_Ligase_1.
 DR PROSITE: PS00697; DNA_LIGASE_A1; FALSE_NEG.
 DR PROSITE: PS00333; DNA_LIGASE_A2; FALSE_NEG.
 DR PROSITE: PS0160; DNA_LIGASE_A3; 1.
 KM DNA repair: DNA replication; DNA recombination; Cell division; Ligase;
 KM ATP-binding: NAD. 249 AMP (BY SIMILARITY).
 FT BINDING 249
 SQ SEQUENCE 559 AA; 63748 MW; 91AB32542E03D20D CRC64;

Query Match 9.9%; Score 70.5; DB 1; Length 559;
 Best Local Similarity 29.7%; Pred. No. 38;
 Matches 22; Conservative 13; Mismatches 30; Indels 9; Gaps 3;

QY 42 PHYVEF-PI-KPGTTLTKKEIYEVWALDPAVKEFRVVELDPSAKIEVYYDKKKKE 99
 DB 456 PHSEGFLEPVKGVSGFDELVEFTKMLKRYIRQEGKFEIEPKVYIEVY-----Q 508
 Db 509 EIQKSPFKSGFAL 522

QY 100 ETKSFPITEKGFV 113
 Db 509 EIQKSPFKSGFAL 522

RESULT 43
 YD6_YEAST STANDARD; PRT; 608 AA.
 AC 004399;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE Putative multicopper oxidase YDR506C (EC 1.-.-.-).
 GN YDR506C OR D9719.12.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI_TaxID=4932;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Dietrich F.S., Mulligan J., Allen E., Arayajo R., Aviles E.,
 RA Berno A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
 RA Huntick-Smith S., Hyman R., Kemp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Oehler P., Oh C., Petel F.X.,
 RA Roberts D., Schramm S., Schroeder M., Shogren T., Shoff N.,
 RA Winant A., Yelton M., Botstein D., Davis R.W.;
 RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: COULD BE A IRON TRANSPORT MULTICOPPER OXIDASE, WHICH IS
 CC REQUIRED FOR FERROUS IRON HIGH AFFINITY UPTAKE. MAY BE REQUIRED TO
 CC OXIDIZE FE(II) AND RELEASE IT FROM THE TRANSPORTER. ESSENTIAL
 CC COMPONENT OF COPPER-DEPENDENT IRON TRANSPORT (BY SIMILARITY).
 CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
 CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
 CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: TO IRON TRANSPORT MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U33057; AAB64948.1; -
 DR SGD: S0002914; YDR506C.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR002355; Multicu_oxidase2.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR PROSITE: PS00079; MULTICOPPER OXIDASE1; 2.
 DR PROSITE: PS00080; MULTICOPPER OXIDASE2; 1.
 KM Hypothetical protein; Transport; Oxidoreductase; Copper; Repeat;
 KM Metal-binding; Iron transport.
 FT DOMAIN 51 163
 FT DOMAIN 243 374 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 421 548 PLASTOCYANIN-LIKE 2.
 FT METAL 100 100 PLASTOCYANIN-LIKE 3.
 FT METAL 102 102 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 102 102 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 145 145 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 147 147 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 147 147 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 452 452 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 452 452 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 455 455 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 457 457 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 530 530 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 531 531 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 532 532 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 536 536 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL ? ? COPPER (TYPE 1) (BY SIMILARITY).
 SQ SEQUENCE 608 AA; 69281 MW; D51B830F15E704AC CRC64;

Query Match 9.9%; Score 70.5; DB 1; Length 608;
 Best Local Similarity 29.0%; Pred. No. 42;
 Matches 29; Conservative 16; Mismatches 38; Indels 17; Gaps 6;

QY 41 SPHYVEFPIKPGTTLTKKEIYEVWALDPAVKEFRVVELDPSA--KIEVYY--DNK 96
 DB 355 NPEDVSIHLPGET-----KAELYRDIEPTQENKLR-TKADPAVAFEDVAYVYDESTK 409
 QY 97 KKEETSPITEKGFVPLDSEHKNP--GPNLTGVY 133
 DB 410 QKYGCMKYNERTF-----SEYKDPVAFGNYETDIYI 444

RESULT 44
 YD6_MYCPN STANDARD; PRT; 115 AA.
 ID YD46_MYCPN
 AC P75432;
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MBN346 (H91_orf115).
 GN MBN346 OR MP490.
 OS Mycoplasma pneumoniae.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Mycoplasmatelaceae; Mycoplasma.
 NC NCB1_taxonomy:2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE-97105885; PubMed-8948633;
 RA Himmelfeld R., Hilbert H., Pliagens H., Pirkl E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AE000048; AAB96138.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 115 AA; 13468 MW; 95E6EBBE9F054B CRC64;
 Query Match 9.8%; Score 70; DB 1; Length 115;
 Best Local Similarity 25.3%; Pred. No. 6.5;
 Matches 25; Conservative 13; Mismatches 31; Indels 30; Gaps 3;
 QY 54 TLTKREIYVEMALD-----ATAYKE-FRVEYELDPSAKIEVYDKKK 97
 DB 29 TLNERIKHYDTDKLIDATQNPKNDEPNPKESGFSNIFACESKEAAIEYKPKK 88
 QY 98 KEETKSPFTEKGFVYVPLSEHKINPGFNLTQVIEKK 136
 DB 89 QIEQKKLPKIKASITIAQ-----TKVYMNKK 113
 RESULT 45
 ID Y255_METVA STANDARD; PRT; 251 AA.
 AC Q57703;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein M0255.
 GN M0255.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 NC NCB1_taxonomy:2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE-96337999; PubMed-8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weissflog K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhlman J.L., Nguyen D.,
 RA Uitterlbeck T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii";
 RL Science 273:1058-1073(1996).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U67480; AAB98242.1; -
 DR TIGR; M0255; -
 DR InterPro: IPR003830; DUF210.
 DR Pfam: PF02679; DUF210.1
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 251 AA; 28368 MW; 9B3F2CD349808DE4C CRC64;
 Query Match 9.8%; Score 70; DB 1; Length 251;
 Best Local Similarity 25.5%; Pred. No. 16;
 Matches 40; Conservative 19; Mismatches 44; Indels 54; Gaps 9;
 QY 29 VTGVDSKGNELSPHYVEFPK-----GVT-----LTKEIYVEMALDA 70
 DB 15 LTVVLDKG---LPPKFYEDLYKVGQYIDVKKMGTSAYIDRDVYKKNYKKGKGINV 71
 QY 71 -----TAYKE-----FRVEL-DPSAKIEVYDKNKKETKSPFI- 106
 DB 72 YPGCTLFYVAYSKGKDFEFLNCEKLGFEVVEISDSSDISERRNAJKRAKNGFVVL 131
 QY 107 TEKGFFVVD-----LSEHKINPGFNLTQVIE 134
 DB 132 TEYKKRPDKDQLTIDRIKLINFDDAGADYVIE 168
 RESULT 46
 ID KFM5_FSVMD STANDARD; PRT; 978 AA.
 AC P00545; O86597;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase transforming protein fms (CC 2.7.1.112).
 GN v-fms.
 OS Feline sarcoma virus (strain McDonough).
 CC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
 NC NCB1_taxonomy:11778;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-84119469; PubMed-6582485;
 RA Hampe A., Gobet M., Sherr C.J., Galibert F.;
 RT "Nucleotide sequence of the feline retroviral oncogene v-fms shows
 RT unexpected homology with oncogenes encoding tyrosine-specific protein
 RT kinases";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:85-89(1984).
 RN [2]
 RP REVISIONS. SEQUENCE FROM N.A.
 RX MEDLINE-92015516; PubMed-1833563;
 RA Smola U., Hennig D., Hadwiger-Fangmeier A., Schuetz B., Pfaff E.,
 RT "Reassessment of the v-fms sequence: threonine phosphorylation of the
 RT COOH-terminal domain";
 RL J. Virol. 65:6181-6187(1991).
 CC -1- FUNCTION: v-fms IS DERIVED FROM THE RECEPTOR FOR COLONY
 CC STIMULATING FACTOR 1 (CSF-1).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-FMS
 CC POLYPROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
 CC PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; K01643; AAA43045.1; -
 DR EMBL; S59588; AAB20028.1; -
 DR PIR; A00654; TVMVM. -
 DR HSSP; P11362; IECK.
 DR InterPro; IPR000719; Euk_DK_kinase.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR003600; Ig_Like.
 DR InterPro; IPR001824; Receptor_tyr_kin_III.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00047; Ig_5.
 DR Pfam; PF00069; Kinase; 2.
 DR SMART; SM00410; Ig_Like; 4.
 DR SMART; SM00408; IGC2; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR Polypeptide; Tyrosine-protein kinase; Oncogene; Transferase; Receptor;
 KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;
 KW Immunoglobulin domain; Repeat.
 FT DOMAIN 1 543 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 544 568 POTENTIAL.
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 569 978 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 58 138 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 141 231 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 238 332 IG-LIKE C2-TYPE DOMAIN 4.
 FT DOMAIN 333 431 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 432 536 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 613 942 PROTEIN KINASE.
 FT NP_BIND 619 627 ATP (BY SIMILARITY).
 FT BINDING 647 647 ATP (BY SIMILARITY).
 FT DISULFID 76 118 POTENTIAL.
 FT DISULFID 161 211 POTENTIAL.
 FT DISULFID 258 312 POTENTIAL.
 FT DISULFID 451 516 POTENTIAL.
 FT MOD_RES 841 841 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 973 973 PHOSPHORYLATION.
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 524 524 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 714 714 L -> P (IN REF. 1).
 FT CONFLICT 971 978 ORTPVAR -> RGPPL (IN REF. 1).
 FT SEQUENCE 978 AA; 108491 MW; 4C7CA4835185EBF CRC64;

Query Match 9.8%; Score 70; DB 1; Length 978;
 Best Local Similarity 27.1%; Pred. No. 83;
 Matches 26; Conservative 18; Mismatches 42; Indels 10; Gaps 5;

OY 19 EPTGP---YLVNVTGVDSKGNELSPRVEPIKPTLTLEKI-EYVEVALDATAVK 74
 DB 33 EAMGPALLVLLMATAMAHAGVPIOPSGPELVYVGGTIVLRGVNGSVEM--DGPISP 90
 OY 75 EPRVVELDSAKIEVYVYKKNKKEETKSPITEKG 110
 DB 91 HNN-LDIDPPSSILTT---NNATFQNTGYHCTEPG 122

RESULT 47
 NXP2_MOUSE STANDARD; PRT; 244 AA.
 ID NXP2_MOUSE
 AC 061199;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Neurexophilin 2 precursor (Fragment).
 GN NXP2 OR NPH2.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_Taxid=10090;

RP SEQUENCE FROM N.A.
 RX MEDLINE=96285495; PubMed=8699246;
 RA Petrenko A.G., Ullrich B., Missler M., Krasnoperov V., Rosahl T.W.,
 RA Suedhof T.C.;
 RT "Structure and evolution of neurexophilin";
 RL J. Neurosci. 16:4360-4369(1996).

CC -1- FUNCTION: MAY BE SIGNALING MOLECULES THAT RESEMBLE NEUREPEPTIDES
 AND THAT ACT BY BINDING TO ALPHA-NEUREXINS AND POSSIBLY OTHER
 RECEPTORS (POTENTIAL).

CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -1- SIMILARITY: BELONGS TO THE NEUREXOPHILIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U56650; AAB18763.1; -
 DR MGD; MGI:107491; Nxp2.

KW Glycoprotein; Signal.
 FT NON_TER 1 1
 FT SIGNAL <1 2

FT CHAIN 3 244 NEUREXOPHILIN 2.
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 244 AA; 28189 MW; 3D3593CD8E3A178 CRC64;

Query Match 9.7%; Score 69.5; DB 1; Length 244;
 Best Local Similarity 24.7%; Pred. No. 18;
 Matches 41; Conservative 22; Mismatches 42; Indels 61; Gaps 11;

OY 7 GKRYK---GDDAYFEP-----TGP-----YLVNVTGVDSKGNELSP-H 43
 DB 86 GKFKMGMGDDSHENITVKLMLITGKIYDHGNGFSPFRHNSGLGNVSLVPPSR 145

OY 44 YVEPIKPTGL-TREKEY--YEMALDATAVEFVVELDSAKIEVYVYKKNKKEE 100
 DB 146 YVEFEISQSTLETKRESFNCHIEKTDRA-KTALCNEDFS--KICY-----OEQ 195

OY 101 TRS-----PPITEKGFVVDLSHIKNP 123
 DB 196 TQSHVSWLCSKPRFYICHIIFYSVDYKLVOK--VCPDYNVHSERP 239

RESULT 48
 GYRB_TREDE STANDARD; PRT; 638 AA.
 ID GYRB_TREDE
 AC 087545;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE DNA gyrase subunit B (EC 5.99.1.3).
 GN GYRB.
 OS Treponema denticola.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=158;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35405;
 RA Greene S.R., Stamm L.V.;
 RT "Characterization of the gyrb region of Treponema denticola."; Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
 RL
 CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining of double-stranded DNA.
 CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE B CHAIN IS RESPONSIBLE FOR DNA BREAKAGE AND REJOINING. THE B CHAIN CATALYZES ATP HYDROLYSIS. THE ENZYME FORMS AN AZB2 TETRAMER.
 CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF083949; AAC62070.1; -
 DR HSSP: P06982; 1A76.
 DR InterPro: IPR002288; DNA_gyraseB_C.
 DR InterPro: IPR001241; DNA_topoisomI.
 DR InterPro: IPR003594; HATPase_c.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF00986; DNA_gyraseB_C_1.
 DR Pfam: PF0204; DNA_topoisomI; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF01751; Toprim; 1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR PRODOM: PD00616; DNA_topoisomI; 1.
 DR PRODOM: PD149633; DNA_gyraseB_C; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00433; TOP2c; 1.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 DR Topoisomerase; Isomerase; ATP-binding.
 KW
 SQ SEQUENCE 638 AA; 71391 MW; A0DDFFA091C2D5A6 CRC64;

 Query Match 9.7%; Score 69.5; DB 1; Length 638;
 Best Local Similarity 23.8%; Pred. No. 56;
 Matches 40; Conservative 24; Mismatches 55; Indels 49; Gaps 9;

 QY 4 FDGKXK-----KGDASYEPTGPLYMNVGV--DSKGNELSPHYEF-----PI 49
 DB 104 FDGSKTKVSGGLHGVGVS-----VNVALSYMMEAVYKDGFEHYAKFKSGSLTEPV 154
 QY 50 KP-GTTLTREKTEIYVEMALDATAYKFRVLEDP-----SAKIEVYYDK--N 95
 DB 155 KRIGEL---EKSGTVIRMAVDPISIFETIYVNEVLAIRLRELAFLNKSISTIMDBERS 211
 QY 96 KKEEERKSP-----ITEKGFVYVPLDSEIRKINPGNLTIKYVIE 134
 DB 212 TPKEVFAFEGGISQFVSYLNEKQVFPKSPVFIIEGKNDILCEVAIQ 259

 RESULT 49
 ID CADI_XENIA STANDARD; PRT; 872 AA.
 AC P30944; Q91709;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Epithelial-cadherin precursor (E-cadherin) (Vvomorulin) (XTCAD-1).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; CC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8353;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95324376; PubMed=7600966;
 RA Levine E., Lee C.H., Kintner C., Gumbiner B.M.;
 RT "Selective disruption of E-cadherin function in early Xenopus embryos by a dominant negative mutant."; Development 120:901-909(1994).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Tail bud;
 RX MEDLINE=94368839; PubMed=8086449;
 RA Tool O., Fujii G., Tashiro K., Shiohara K.;
 RT "Molecular cloning of cDNA for XTCAD-1, a novel Xenopus cadherin, and its expression in adult tissues and embryos of Xenopus laevis."; Biochim. Biophys. Acta 1219:121-128(1994).
 RL [3]
 RN SEQUENCE OF 149-872 FROM N.A.
 RX MEDLINE=94363396; PubMed=8081882;
 RA Broders F., Girault J.M., Simonneau L., Thierly J.P.;
 RT "Sequence and distribution of Xenopus laevis E-cadherin transcripts."; Cell Adhes. Commun. 1:265-277(1993).
 RN [4]
 RP SEQUENCE OF 149-169.
 RX MEDLINE=91347911; PubMed=1879345;
 RA Angres B., Mueller A.H.J., Kellermann J., Hansen P.;
 RT "Differential expression of two cadherins in Xenopus laevis."; Development 111:829-844(1991).
 RL
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS. THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC MANNER IN CONNECTING CELLS. CADHERINS MAY THIS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN IS A LIGAND FOR INTEGRIN ALPHA-E/BETA-7.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.
 CC -1- DEVELOPMENTAL STAGE: APPEARS IN THE EMBRYONIC ECTODERM DURING GASTRULATION WHEN EPIDERMAL DIFFERENTIATION COMMENCES AND IT DISAPPEARS FROM THE NEURAL PLATE AREA UPON NEURAL INDUCTION.
 CC -1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U04708; AAA93116.1; -
 DR EMBL: L29057; AAA61489.1; -
 DR EMBL: X75454; CAA53206.1; -
 DR PIR: A60128; A60128.
 DR HSSP: P09803; 1SUH.
 DR InterPro: IPR002126; Cadherin.
 DR InterPro: IPR000233; Cadherin_C_term.
 DR Pfam: PF00028; cadherin; 5.
 DR PRINTS: PR01049; Cadherin_C_term; 1.
 DR PRINTS: PR00205; CADHERIN.
 DR SMART: SM00112; CA; 4.
 DR PROSITE: PS00232; CADHERIN_1; 3.
 DR PROSITE: PS50268; CADHERIN_2; 4.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat; Signal.
 KW
 FT SIGNAL 1 25 POTENTIAL.
 FT PROPEP 26 148
 FT CHAIN 149 872 EPITHELIAL-CADHERIN.

```

FT DOMAIN 149 701 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 702 722 POTENTIAL.
FT DOMAIN 723 872 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 244 254 CADHERIN 1.
FT DOMAIN 358 368 CADHERIN 2.
FT DOMAIN 577 587 CADHERIN 3.
FT DOMAIN 710 721 POLY-LED.
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 242 242 I -> V (IN REF. 3).
FT CONFLICT 332 332 N -> T (IN REF. 3).
FT CONFLICT 487 487 E -> R (IN REF. 3).
FT CONFLICT 501 502 AT -> CS (IN REF. 3).
FT CONFLICT 539 541 GNG -> EMA (IN REF. 3).
FT CONFLICT 548 548 K -> R (IN REF. 3).
FT CONFLICT 557 557 V -> G (IN REF. 3).
FT CONFLICT 567 567 P -> L (IN REF. 3).
FT CONFLICT 602 604 GFR -> EPQ (IN REF. 3).
FT CONFLICT 638 640 GOS -> DK (IN REF. 2).
FT CONFLICT 640 647 SILELRPK -> VYLSDL (IN REF. 3).
FT CONFLICT 660 660 T -> A (IN REF. 3).
FT CONFLICT 770 770 A -> S (IN REF. 3).
FT CONFLICT 842 842 P -> S (IN REF. 2 AND 3).
FT CONFLICT 871 871 DE -> GED (IN REF. 3).
SQ SEQUENCE 872 AA; 96065 MM; 08637967EEFB4664 CRC64;

```

```

Query Match 9.7%; Score 69.5; DB 1; Length 872;
Best Local Similarity 25.5%; Pred. No. 81;
Matches 28; Conservative 17; Mismatches 46; Indels 19; Gaps 4;

```

```

OY 8 KYKKGDDASYE---PTGPR-LMVNVGVDSKGNELSPHY-----EPIKRG 52
DB 409 KIKKNEANYSIOTDTGNIGLTKVKGDIY---LKKOILSIYTNKANSFPLDTS 464
OY 53 TLTREKIEYVEMALDATAYKEFRVVELDPSAKIEVTVYDNKKKKEETK 102
DB 465 FATVTVSEVDVNEAPIFLPVPYKEVSSEDLPSGVAVATYTTQDDPKDQNO 514

```

```

RESULT 50
ACON_LEGPN STANDARD: PRT; 891 AA.
AC P37032:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Aconitase hydrolase (EC 4.2.1.3) (Citrate hydro-lyase) (Aconitase)
DE (Major iron-containing protein) (MCP) (IP210).
GN ACON
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN-PHILADELPHIA 1;
RX MEDLINE=93374864; PubMed=8366052;
RA Mengaud J.M., Horwitz M.A.;
RT "The major iron-containing protein of Legionella pneumophila is an
RT aconitase homologous with the human iron-responsive element-binding
RT protein."
RL J. Bacteriol. 175:5666-5676(1993).
CC -1- CATALYTIC ACTIVITY: Citrate -> cis-aconitate + H(2)O.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- MISCELLANEOUS: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE
CC (3FE-4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE
CC CATALYTIC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND
CC ISOCITRATE.

```

```

CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: I23081; AAA25295.1; .
DR PIR: B48642; B48642.
DR HSSP: P16276; 7ACN.
DR InterPro: IPR001030; Aconitase.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00330; aconitase; 1.
DR Pfam: PF00694; aconitase_C; 1.
DR PRINTS: PR00415; ACONITASE.
DR PRODOM: PD000511; Aconitase; 1.
DR PROSITE: PS00450; ACONITASE_1; 1.
DR PROSITE: PS01244; ACONITASE_2; 1.
KW Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S.
FT METAL 435 435 IRON (IRON-SULFUR CLUSTER)
FT METAL 501 501 (BY SIMILARITY).
FT METAL 501 501 IRON (IRON-SULFUR CLUSTER)
FT METAL 504 504 (BY SIMILARITY).
FT METAL 504 504 IRON (IRON-SULFUR CLUSTER)
FT METAL 504 504 (BY SIMILARITY).
SQ SEQUENCE 891 AA; 98209 MM; F81FC99520859B48 CRC64;

```

```

Query Match 9.7%; Score 69.5; DB 1; Length 891;
Best Local Similarity 19.3%; Pred. No. 83;
Matches 41; Conservative 25; Mismatches 51; Indels 95; Gaps 8;

```

```

OY 7 KYKKGDDASYEPTGYLVAVNT-----GVDSKGNELSPHYEPIKGTTLT----- 56
DB 258 GKLEKG-----TTARDLVLTVMQDKRGVYKGFYFPGGLNDPLDRAVTISMAP 310
OY 57 -----KKIEYVEMALDATAYKEFRVVELDPSAKIEVTVYDNK----- 95
DB 311 EYGATGFFPYDKETIK-----LELTGRDKHTALVEAYAKAGMYDDNDNEPFTDS 365
OY 96 -----KKKEYSFPIYKGF----- 111
DB 366 LHLDLGVSPESLAGPKRPQKVNLSLPVEFNFLIEVGKEKEKEKFAVKNKDFQMKHG 425
OY 112 --VVPDLSE--HKNPGF-----NITRVVIEK 135
DB 426 HVIYIAITSTCTNTSPSVLMAAGLVAKKALEK 457

```

Search completed: September 1, 2002, 03:27:20
Job time: 511 sec

...

Gencore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2002, 03:17:47 ; Search time 69.56 Seconds

(without alignments)
338,230 Million cell updates/sec

Title: US-09-728-670-10

Perfect score: 715
Sequence: 1 SSSFDKGYKKKGDDASYEP.....SEHKNPGFNITKVIIEKK 136

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	163	16	099SU7 staphylococ
2	710	99.3	163	2	09LC46 staphylococ
3	703	98.3	163	3	09AM04 staphylococ
4	699	97.8	136	2	033929 staphylococ
5	228	31.9	377	2	0934E7 streptococ
6	138.5	19.4	416	2	092FE2 streptococ
7	88	12.3	569	10	092R86 zea mays (m
8	87	12.2	348	16	034353 bacillus su
9	84.5	11.8	483	9	09G004 bacterioph
10	83.5	11.7	606	16	09CMA9 pasteurella
11	83	11.6	611	16	09X0Y4 thermotoga
12	82.5	11.5	16215	5	09NFS3 drosophila
13	81	11.3	526	2	09RO87 mycoplasma
14	80.5	11.3	158	17	059474 pyrococcus
15	80	11.2	543	17	0971U1 pyrolobus
16	79.5	11.1	378	5	0950B3 caenorhabdi

17	78	10.9	350	10	039980 hyoseyamus
18	77.5	10.8	186	10	09LPB1 arabidopsis
19	77	10.8	559	10	065323 capsicum an
20	77	10.8	559	10	081923 capsicum an
21	77	10.8	608	11	099J21 mus musculu
22	76.5	10.7	289	17	029463 archeoglob
23	76.5	10.7	349	17	0971G3 sulfolobus
24	76	10.6	252	16	09CEA0 lactococcus
25	75	10.5	933	11	091XY8 mus musculu
26	74.5	10.4	666	2	P96018 staphylococ
27	74	10.3	233	5	09X002 caenorhabdi
28	74	10.3	233	17	029007 archeoglob
29	74	10.3	236	5	09X0Q3 caenorhabdi
30	74	10.3	462	10	004622 arabidopsis
31	74	10.3	480	10	09M156 arabidopsis
32	74	10.3	552	16	09PEW0 ureaplasma
33	73.5	10.3	257	16	092N17 rhizobium m
34	73.5	10.3	431	10	09X1R2 arabidopsis
35	73.5	10.3	764	3	042976 schizosach
36	73.5	10.3	792	11	091X25 mus musculu
37	73.5	10.3	1536	16	09X079 thermotoga
38	73	10.2	171	11	09CUN9 mus musculu
39	73	10.2	345	13	091946 gallus gall
40	73	10.2	419	17	0973T1 sulfolobus
41	73	10.2	470	10	022732 arabidopsis
42	73	10.2	654	4	09ERF0 oryza sativ
43	73	10.2	875	4	09H706 homo sapien
44	73	10.2	2505	11	063577 rattus norv
45	73	10.2	2509	4	Q16702 homo sapien

ALIGNMENTS

RESULT 1
ID Q99SU7 PRELIMINARY; PRT; 163 AA.
AC Q99SU7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE STAPHYLOKINASE PRECURSOR.
GN SAK OR SAK1758.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacilli; Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158679;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Iino T., Kanemori M.,
RA Matsunari H., Maruyama A., Murakami H., Hoshiyama A., Mizutani-Uji Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus".
RT Lancet 357:1225-1240(2001).
RL EMBL: AP003135; BAB43032.1; -.
DR HSSP: P00802; ISSN.
DR InterPro: IPR004093; Staphylokinase.
DR Pfam: PF02821; Staphylokinase; 1.
KW Complete proteome.
SQ SEQUENCE 163 AA; 18520 MW; 413CDAE14BE9FD40 CRC64;

Query Match 100.0%; Score 715; DB 16; Length 163;
Best Local Similarity 100.0%; Pred. No. 3.4e-59;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSSFDKGYKKKGDDASYEPFGPIAMVNTGVDSKGNELLSPHYVEPPIKPGITTKREKI 60

```

DB 28 SSSFDGKRYKKGDDASYFEPTGPLYLVNTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 87
QY 61 EYVEWALDATALAKKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFFVVDLSEHI 120
DB 88 EYVEWALDATALAKKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFFVVDLSEHI 147
QY 121 KNPGFNLITKRVIEKK 136
DB 148 KNPGFNLITKRVIEKK 163

```

RESULT 2

```

ID 09LC46 PRELIMINARY: PRT: 163 AA.
AC 09LC46:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE STAPHYLOKINASE.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20219005; PubMed=10754251;
RA Horii T., Yokoyama K., Barua S., Odagiri T., Futamura N., Hasegawa T.,
RA Ohta M.;
RT "The staphylokinase gene is located in the structural gene encoding N-
RT acetyltransferase-L-alanine amidase in methicillin-resistant
RT Staphylococcus aureus."
RL EMBL; AB033232; BAA95011.1; -.
DR HSSP; P00802; ISSN.
DR InterPro; IPR004093; Staphylokinase.
DR Pfam; PF02821; Staphylokinase; 1.
KW kinase.
SQ SEQUENCE 163 AA: 18474 MW: 3DB45E35046029DD CRC64;

```

Query Match 99.3%; Score 710; DB 2; Length 163;
 Best Local Similarity 98.5%; Pred. No. 9.9e-59;
 Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 SSSFDGKRYKKGDDASYFEPTGPLYLVNTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 60
DB 28 SSSFDGKRYKKGDDASYFEPTGPLYLVNTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 87
QY 61 EYVEWALDATALAKKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFFVVDLSEHI 120
DB 88 EYVEWALDATALAKKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFFVVDLSEHI 147
QY 121 KNPGFNLITKRVIEKK 136
DB 148 KNPGFNLITKRVIEKK 163

```

RESULT 3

```

ID 09AM04 PRELIMINARY: PRT: 163 AA.
AC 09AM04:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE STAPHYLOKINASE SAKXH.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RA Wei W., Xiang H., Tan H.;

```

```

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332619; AA011636.1; -.
DR HSSP; P00802; ISSN.
KW Kinase.
SQ SEQUENCE 163 AA: 18509 MW: 0A266B5DEAF65B5A CRC64;

```

Query Match 98.3%; Score 703; DB 2; Length 163;
 Best Local Similarity 97.8%; Pred. No. 4.4e-58;
 Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 SSSFDGKRYKKGDDASYFEPTGPLYLVNTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 60
DB 28 SSSFDGKRYKKGDDASYFEPTGPLYLVNTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 87
QY 61 EYVEWALDATALAKKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFFVVDLSEHI 120
DB 88 EYVEWALDATALAKKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFFVVDLSEHI 147
QY 121 KNPGFNLITKRVIEKK 136
DB 148 KNPGFNLITKRVIEKK 163

```

RESULT 4

```

ID 033929 PRELIMINARY: PRT: 136 AA.
AC 033929:
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE STAPHYLOKINASE (FRAGMENT).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 29213;
RX MEDLINE=97417758; PubMed=9271816;
RA Kim S.H., Chun H.S., Han M.H., Park N.Y., Suk K.;
RT "A novel variant of staphylokinase gene from Staphylococcus aureus
RT ATCC 29213."
RL Thromb. Res. 87:387-395(1997).
DR EMBL; U77328; AAB84174.1; -.
DR HSSP; P00802; ISSN.
DR InterPro; IPR004093; Staphylokinase.
DR Pfam; PF02821; Staphylokinase; 1.
KW Kinase.
FT NON-TER
SQ SEQUENCE 136 AA: 15551 MW: 79916B136CA1A3F CRC64;

```

Query Match 97.8%; Score 699; DB 2; Length 136;
 Best Local Similarity 97.8%; Pred. No. 8.5e-58;
 Matches 133; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 SSSFDGKRYKKGDDASYFEPTGPLYLVNTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 60
DB 1 SSSFDGKRYKKGDDASYFEPTGPLYLVNTGVDSKGNELSPHYVEFPPIKPGTTLTREKI 60
QY 61 EYVEWALDATALAKKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFFVVDLSEHI 120
DB 61 EYVEWALDATALAKKEFRVVELDPSAKIEVYYDKNKKKEETKSFPITEKGFFVVDLSEHI 120
QY 121 KNPGFNLITKRVIEKK 136
DB 121 KNPGFNLITKRVIEKK 136

```

```

RESULT 5
O934F7 PRELIMINARY: PRT: 377 AA.
ID O934F7

```


AC		O334F7;		
DT	01-DEC-2001	(TREMBLrel. 19, Created)		
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	PLASMINOGEN ACTIVATOR PRECURSOR.			
GN	PAUB.			
OS	Streptococcus uberis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OX	Streptococcus.			
OX	NCBI_TaxID=1349;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-SK880:			
RA	Ward P.N., Leigh J.A.;			
RT	*Characterisation of PAUB, a novel broad spectrum plasminogen activator from Streptococcus uberis.*;			
RL	J. Bacteriol. 0:0-0(0).			
DR	EMBL; AJ314852; CAC85651.1; -.			
KW	Signal.			
FT	SIGNAL	1	25	POTENTIAL.
FT	CHAIN	26	377	PLASMINOGEN ACTIVATOR.
SO	SEQUENCE	377 AA;	43841 MW;	9283CBAA57EF5E4F CRC64;

	Query Match	31.9%; Score 228; DB 2; Length 377;
	Best Local Similarity	35.8%; Pred. No. 2,5e-13;
	Matches	44; Conservative 30; Mismatches 43; Indels 6; Gaps 1.
OY	14 DASFEPTGPLYLVNVTGVDSKGNELLSPHYEPPIKPGTLLTKERIEIYEMALDATAY	73
	: :	
Dd	30 EVNGDEPQGTMTLTINITGSKEGQLLSPREVMQFRLTAGEVINKNELLDKVGLVIDSAAS	89
OY	74 KEFRVVELDPSKKILEVIYYDDKNKKKEEIKSFPIITEKGFPVPLSEHIKPKGFNLITKYVI	133
	: :	
Dd	90 NOFEVVDKRPESKVEMSYDFDRHLELD-----ITERGEFIVPDIYSYEKKPSFLLTGPVII	143
OY	134 EKK 136	
	: : :	
Dd	144 QRK 146	

RESULT	6			
092FE2				
ID	Q92FE2	PRELIMINARY;	PRT;	416 AA.
AC	Q92FE2;			
DT	01-MAY-1999 (TREMBLrel.. 10, Created)			
DT	01-MAY-1999 (TREMBLrel.. 10, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel.. 19, Last annotation update)			
DE	STREPTOKINASE (FRAGMENT).			
GN	SKC.			
OS	Streptococcus equisimilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
CC	Streptococcus.			
OX	NCBI_TaxID=119602;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=87-542-W;			
RX	MEDLINE=20038313; PubMed=10569766;			
RA	Cadallero A.R., Lottenberg R., Johnston K.H.;			
RT	*Cloning, expression, sequence analysis, and characterization of			
RT	streptokinases secreted by porcine and equine isolates of			
RT	streptococcus equisimilis".			
RL	Infect. Immun. 67:6478-6486(1999).			
DR	EMBL; AF104301; M8C83933.1; "			
DR	InterPro; IPR004093; staphylokinase.			
DR	Pfam; PF02821; staphylokinase; 1.			
FT	NON_TER	416	416	
SO	SEQUENCE	416 AA;	47869 MW;	CAB79D18496BCCC4 CRC64;

Query Match	19.48;	Score 138.5;	DB 2;	Length 416;
Best Local Similarity	28.18;	Pred. No. 6.3e-05;		

	Matches	38;	Conservative	27;	Mismatches	49;	Indels	21;	Gaps	4;
QY	7	GATKKKGDA----	STFEPTGPLYMANVTG	VDKSGNELLSHYVEPIK	GGTLLTEKIE	61				
		:	: :	:	: :					
Db	47	GTVQPTDPEFNKSDQ	PTGTLGLIININVG	VDSRHNTLVLPO	YODIKLDNKTKYTD	106				
		:	: :	:	:					
QY	62	YVYEMALDAAHYKEP	RVVELDPSAKIEVTY	YD-----	KNKKKEETKSP	PTTEKG	110			
		:	: :	:	:					
Db	107	RFVQNVVIDA-SMNF	KVYIDLHEVLKVT	RTATQDEMSPT	FTSLAENQNDAN	OSYLLTG	165			
		:	: :	:	:					
QY	111	FVVP----	DLSEHIK	121						
		:	: :	:	:					
Db	166	LVEPVS	KRNLEK	EVYIK	180					
		:	: :	:	:					

RESULT	7			
09ZR86				
ID	09ZR86	PRELIMINARY;	PRT;	569 AA.
AC	09ZR86;			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	PII-LIKE PROTEIN.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;			
OC	Panicoidae; Andropogoneae; Zea.			
OX	NCBI_TaxID=4577;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99026300; PubMed=9808743;			
RA	Laugner B.J., Sehne P.C., Feil R.J.;			
RT	"A novel nuclear member of the thioredoxin superfamily.";			
RL	Plant Physiol. 118:987-996(1998).			
	[2]			
RP	SEQUENCE FROM N.A.			
RA	Laugner B., Feil R.;			
RL	Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: U90944; AAD04231.1; -			
DR	InterPro: IPR004146; DC1.			
DR	InterPro: IPR000063; Thired.			
DR	Pfam: PF03107; DC1; 1.			
DR	Pfam: PF00085; Thired; 2.			
DR	PROSITE: PS00194; THIOREDOXIN; UNKNOWN_1			
SO	SEQUENCE 569 AA; 63356 MW; 5472A399IDCE3300 CRC64;			

	Query Match	12.3%; Score 88;	DB 10;	Length 569;
	Best Local Similarity	28.4%;	Pred. No. 4;7;	
Matches	31;	Conservative	19;	Mismatches 39; Indels 20; Gaps 5
OY	23 PYLWANTGYDSGNNLLS-----PHYVEPIKPTTLTKRIEYYEMALDATAYK 74	: : : :	:	:
Db	97 PWLAVPSS--DSDGRERALDGOFVSGCIPHLVLTDLKTEEVYTEDEGVSEYGEAVPFT 154	: :: :: :	::: :: :	:
OY	75 EFRVVELDPQAKIEVTYYDNKKKE----EKSKPRKEGFVPDLSE 118	: : :	: :: :: :	:
Dd	155 PDRINELKEOEKAEE---KENOTIQSVLTSTRDYLINSKGDYKP-ISE 198	: :: :: :	:	:
RESULT	8			
ID	034353	PRELIMINARY:	PRT:	348 AA.
AC	034353:			
DT	01-JAN-1998 (TrEMBLrel. 05,	Created)		
DT	01-JAN-1998 (TrEMBLrel. 05,	Last sequence update)		
DT	01-NOV-1998 (TrEMBLrel. 08,	Last annotation update)		
DE	YDJN PROTEIN.			
GJ	YDJN.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
CC	Bacillus/Staphylococcus group; Bacillus.			
NCBI_taxid=1423;				

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MARBURG 168;
RA Kaashara Y., Nakai S., Ogasawara N., Yata K., Sadale Y.;
RL Dna Res. 0:0-0(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Barbero M.G., Besleres P., Bolotin A., Borchert S.,
RA Borriest R., Bourcier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brovelli S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fadret C., Ferrari E., Foulger D.,
RA Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Katamata D., Kaashara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Moore D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Preecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowaka A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpiltra P., Tognoni A.,
RA Toseito V., Uchiyama S., Vandenbol M., Vannier F., Vassacott A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Westenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007638; BAA22769.1; -
DR EMBL; 299107; CAB12445.1; -
KW Complete proteome.
SQ SEQUENCE 348 AA; 39710 MW; F297D69DAC72967D CRC64;

Query Match 12.2%; Score 87; DB 16; Length 348;
Best Local Similarity 26.5%; Pred. No. 3.2;
Matches 31; Conservative 18; Mismatches 34; Indels 34; Gaps 5;

OY 5 DKGKVKGGDADYFEPTGYLWVNTGVDSKGNELLSPHYVFPKIPGTLTKKEIEYV 64
DB 27 DKG-HEKADVSNTESGDELLVSTDTD-----LTKFY-----ENDKVIHEELTISYP 75

OY 65 EWALDATAYKEFRVVELDPSAKIEVYVDKNNKKEETKSFPITKEGFVVDLSEHIK 121
DB 76 AFALD-----OKQOVIYTYGNNEQNMRLFKL-----DLKSHK 109

RESULT 9
OY 09G004 PRELIMINARY; PRT; 483 AA.
AC 09G004;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ORF1.
DE bacteriophage phi ETM.

```

```

OC Viruses.
OX NCBI_TaxID=106284;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugai M., Yamaguchi T., Hayashi T., Nakasone K., Takami H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20566787; PubMed=1115106;
RA Yamaguchi T., Hayashi T., Takami H., Nakasone K., Ohnishi M.,
RA Nakayama K., Yamada S., Komatsuzawa H., Sugai M.;
RT "Phage conversion of exfoliative toxin A production in Staphylococcus
RT aureus."
RL Mol. Microbiol. 38:694-705(2000).
DR EMBL; AP001553; BAA97627.1; -
SQ SEQUENCE 483 AA; 55712 MW; DDBB2547E0DC80 CRC64;

Query Match 11.8%; Score 84.5; DB 9; Length 483;
Best Local Similarity 25.5%; Pred. No. 8.2;
Matches 27; Conservative 21; Mismatches 27; Indels 31; Gaps 5;

OY 29 VTGVDSKGNELLSPHYE-----PIKRG-----TTLTKKEIEYVEALDATYKKE 76
DB 139 LTGASNGIEMLHPYLDGESEFKLFVPAEGIPWTOKHEHELEAFI----- 186

OY 77 RVVELDPSAKIEVYVDKNNKKEETKSFPITKEGFVVDLSEHIK 122
DB 187 RMYLLEMETKVE--YMDK-----VTVMYYEENGSLIPDYSNNLEN 225

RESULT 10
OY 09CMA9 PRELIMINARY; PRT; 606 AA.
AC 09CMA9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FADD.
DE FADD_2 OR PM0925.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006132; AAK03009.1; -
DR InterPro; IPR000873; AMP-bind.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00455; AMP_BINDING; 1.
KW Complete proteome.
SQ SEQUENCE 606 AA; 69594 MW; FB50E949401E5224 CRC64;

Query Match 11.7%; Score 83.5; DB 16; Length 606;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 38; Conservative 28; Mismatches 62; Indels 35; Gaps 8;

OY 1 SSSFDKGRYKKKDDASYEPTGP-YLWVNTGVDSKGN-ELLSPHYVEFPKIPGTLTKKE 58
DB 434 ADSFTEGDFLKTGDAGEDDPGNLYITDRIKELMKTSNGKYIAPQYLE-----TKGKD 487

OY 59 KIEYVEALDATAYKEFRVY-----ELDPSAK-IEVYTYK-----NK 96
DB 488 KFIQIVADIADAKRYVALIVPCFNSLEEVAKOLNIKYHRLLEIKHSIDILQFEQRI 547

OY 97 KKEETKSFPITKEGFVVDL-----SEHIKNGPFLTKKYIEK 135

```

Db 548 LQKLPSEQIKKFTLLPOAFTTMEEL-PTLKLRRVILIER 589

RESULT 11

09X0Y4 PRELIMINARY; PRT: 611 AA.

AC 09X0Y4; 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.

OS Thermotoga maritima.

OC Bacteria; Thermotogales; Thermotoga.

OX NCBI_TaxID=2336;

RP SEQUENCE FROM N.A.

RC STRAIN=MSB8 / DSM 3109;

RA MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,

RA Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,

RA McDonald L., Utterback T.R., Malek J.A., Ianher K.D., Garrett M.M.,

RA Stewart A.M., Sutton M.D., Pratt M.S., Phillips C.A., Richardson D.,

RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,

RA Salzberg S.L., Smith R.O., Venter J.C., Fraser C.M.;

RT *Evidence for lateral gene transfer between Archaea and Bacteria from

RT genome sequence of Thermotoga maritima.;

RL Nature 393:323-329(1999).

DR EMBL: AE001781; AAD36331.1; -

DR TIGR: TM1257; -

KW Complete proteome.

SO SEQUENCE 611 AA; 70010 MW; 7FB042DAE49097AB CRC64;

Query Match 11.6%; Score 83; DB 16; Length 611;

Best Local Similarity 21.5%; Pred. No. 15; Matches 28; Conservative 27; Mismatches 45; Indels 30; Gaps 5;

QY 8 KYKKGDDASYEPFG-----PYLAVNVTVGVDKSGN-----ELLSPHYVEFPPIKP 51
 DB 163 RYLLDNGNHINVSQGSVAKTSYTFVLVKSMTSKNDGLMRELRKARITIFNNK 222
 QY 52 GTTTLKREIKIYEVAMATATAYKE-FRYVELDPSAKIYTYVDKKNKKEETKSPITEKG 110
 DB 223 ESLLFLDLRIS--KEWYSERKEMDEMYRVLGIEPKPEFENVARY-----APSRKKG 269
 QY 111 FVVPDLSEHI 120
 DB 270 AYIPDVNKRLL 279

RESULT 12

09NFS3 PRELIMINARY; PRT: 16215 AA.

AC 09NFS3; 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE D-TITIN.

GN SLS OR D-TITIN OR CG1915.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Aphidrodidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Zhang Y.Q., Broadie K.S.;

RT *Characterization of Drosophila D-Titin gene.;

RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AJ271740; CAB93524.1; -

DR HSSP: P56276; ITRK

DR Flybase: FBgn0003432; sls.

DR InterPro: IPR002106; AA_tRNA_ligase_II.

DR InterPro: IPR003961; FN_III.

DR InterPro: IPR003598; Iq_c2.

DR InterPro: IPR003600; Iq_1like.

DR InterPro: IPR003006; Iq_MHC.

DR InterPro: IPR01452; SH3.

DR Pfam: PF00041; fn3_5.

DR Pfam: PF00047; Iq_48.

DR Pfam: PF00018; SH3_1.

DR SMART: SM00408; IGC2_15.

DR SMART: SM00410; Iq_1like; 34.

DR SMART: SM00326; SH3_1.

DR PROSITE: PS00179; AA_tRNA_LIGASE_II_1; UNKNOWN_1.

DR PROSITE: PS50002; SH3_1.

DR Immunoglobulin domain.

SO SEQUENCE 16215 AA; 1841509 MW; 242C8765E00F7603 CRC64;

Query Match 11.5%; Score 82.5; DB 5; Length 16215;

Best Local Similarity 25.5%; Pred. No. 8,2e+02; Matches 36; Conservative 32; Mismatches 52; Indels 21; Gaps 8;

QY 8 KYKKGDDASYFE-----PTGPIYLVNV-TVGVDSKG-NELLSPHYVEFPPIKPTTLK 57
 DB 10444 RHEKGDVQVLEVESVADGEPEEINVISANTEGSEELTDKIK---KRSRKIK 10500
 QY 58 EKIEYVEMALDPAKYEFVVELDPSAKIEVTVYDKNKKEETKS-FPITEKGFVVP-D 115
 DB 10501 DDDDAVYIQQLINA----EIVTELEKYEKIDVQKAKKPKLAKTKRPIDEGETLVG 10556
 QY 116 LSEHKNPFPNLITKVYIEKK 136
 DB 10557 VTEH-EPTKKLTKRKPEKK 10575

RESULT 13

09R0S7 PRELIMINARY; PRT: 526 AA.

AC 09R0S7; 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE MAJOR LIPOPROTEIN.

GN LPPA.

OS Mycoplasma mycoides subsp. mycoides LC.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;

OC Entomoplasmataceae.

OX NCBI_TaxID=44100;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Y-GOAT;

RA MEDLINE=99167669; PubMed=1006658;

RT Monnerat M.P., Thiaucourt F., Poveda J.B., Nicolet J., Frey J.;

RT *Genetic and serological analysis of lipoprotein Lppa in Mycoplasma

RT mycoides subsp. mycoides LC and Mycoplasma mycoides subsp. capri.;

RL Clin. Diagn. Lab. Immunol. 6:224-230(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Y-GOAT;

RA MEDLINE=99440592; PubMed=10512041;

RT Monnerat M.P., Thiaucourt F., Nicolet J., Frey J.;

RT *Comparative analysis of the lppa locus in Mycoplasma capricolum

RT subsp. capricolum and Mycoplasma capricolum subsp. capripneumoniae.;

RL Vet. Microbiol. 69:157-172(1999).

DR EMBL: AF072714; AAF06069.1; -

DR Lipoprotein.

SO SEQUENCE 526 AA; 60288 MW; 60SCEDB5B652249D CRC64;

Query Match 11.3%; Score 81; DB 2; Length 526;

Best Local Similarity 23.7%; Pred. No. 19; Matches 41; Conservative 24; Mismatches 64; Indels 44; Gaps 6;


```

Best Local Similarity 29.7%; Pred. No. 18;
Matches 44; Conservative 17; Mismatches 50; Indels 37; Gaps 10;

OY 14 DASFEPTGPIYMWN-----VTGVDKSG-----NELSPHYVEP---IKPGTITKE 58
DB 219 DAIFF-PSGDITDINDINSYNGSDSKQISSSEDINLRFKPSEREREKSVILLPMATLQMD 277
OY 59 KIEYYEMALDATAKEFRVEL-DPSAKI-----EVTYDKNKKKEETKSPFIT 107
DB 278 IIEVL---ALFTMLMPTGLIELSDCEIKIGDKIKQVLELDPFMYMNAKVEE---PLV 331
OY 108 EKGFTV---PDLSHINKPGFNL-ITRV 131
DB 332 RIGTLVNLIPDVHKSVRRIODDEMTKY 359

RESULT 17
OY 039980 PRELIMINARY; PRT; 350 AA.
AC 039980;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VERISIRADIENE SYNTHASE (FRAGMENT).
OS Hyoscyamus muticus (Egyprian henbane).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Hyoscyamus.
OX NCBI_TaxID=35626;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95221394; PubMed=7706281;
RA Back K., Chappell J.;
RT Cloning and bacterial expression of a sesquiterpene cyclase from
RT Hyoscyamus muticus and its molecular comparison to related terpene
RT cyclases.
RL J. Biol. Chem. 270:7375-7381(1995).
DR HSEBL; U20190; AAA86339.1; -.
DR HSPB; Q40577; SEAU.
DR InterPro: IPR001906; Terp_synth_fam.
DR Pfam: PF01397; Terpene_synth. 1.
FT NON_TER 1
SQ SEQUENCE 350 AA; 41104 MW; FEE71BC7EFDADAE CRC64;

Query Match 10.9%; Score 78; DB 10; Length 350;
Best Local Similarity 32.8%; Pred. No. 23;
Matches 22; Conservative 14; Mismatches 23; Indels 8; Gaps 4;

OY 59 KIEYYEMALDATAKEFRVELDPSAKIEVTYDKNKKKEETKSPFITKEGFV---VPD 115
DB 138 KISKYV--LLD--LYKDYE-TELSKDGKSEVYHAKEMKEIVANYEAKWFLGGMPP 192
OY 116 LSEHIKN 122
DB 193 VSEYLIYN 199

RESULT 18
OY 09LPB1 PRELIMINARY; PRT; 1586 AA.
AC 09LPB1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE T3ZE20.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Shih P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bel Q., Chin C., Chou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T3ZE20 from chromosome
RT 1."
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Chou K., Shih P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bel Q., Chin C., Chou J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC020646; AAF79809.1; -.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR RNA-directed DNA polymerase.
SQ SEQUENCE 1586 AA; 180875 MW; A9D3917EAFDAD790 CRC64;

Query Match 10.8%; Score 77.5; DB 10; Length 1586;
Best Local Similarity 30.4%; Pred. No. 1.5e+02;
Matches 31; Conservative 14; Mismatches 36; Indels 21; Gaps 5;

OY 34 SKNELSPHYVEPIKPGTT-----LTKEIYYEMAL----DATAKEFR 77
DB 1412 AKLNDLMAVYRTAFKTPGTPFNLLYGRSCHLPYE-LEKAMAMVKLNFIDIKTAEEKR 1470
OY 78 VVELDPSAKIEVTYDKNK-KKEETKSF---PTEKGFVPPD 115
DB 1471 LIQLNDLNLKIRLEAYESKIKERTKSFHDKKIVSHDFVGD 1512

RESULT 19
OY 065323 PRELIMINARY; PRT; 559 AA.
AC 065323;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SESQUITERPENE CYCLASE.
OS Capsicum annuum (bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RX MEDLINE=99033462; PubMed=9816674;
RA Back K., He S., Kim K.U., Shih D.H.;
RT Cloning and bacterial expression of sesquiterpene cyclase, a key
RT branch point enzyme for the synthesis of sesquiterpenoid phytoalexin
RT capsidol in UV-challenged leaves of Capsicum annuum.
RL Plant Cell Physiol. 39:899-904(1998).
DR EMBL; AF061285; AAC61260.1; -.

```

DR HSSP: Q40577; SEAT.
DR InterPro: IPR001906; Terp_synth_fam.
DR Pfam: PF01397; Terpene_synth; 1.
SQ SEQUENCE 559 AA; 65098 MW; DBF07B3E3A26DC0 CRC64;

Query Match 10.8%; Score 77; DB 10; Length 559;
Best Local Similarity 34.0%; Pred. No. 49;
Matches 18; Conservative 11; Mismatches 20; Indels 4; Gaps 2;

OY 73 YKFEFVLPDSAKIEVYDKNKKKEKSPITEKGFV---VPDSEHKN 122
DB 357 YKDYE-KEMSRDGRSHVYVYAKERLKLKLVKSYNIEAKWFIEGHMPASEYLKN 408

RESULT 20
081923 PRELIMINARY; PRT; 559 AA.

AC 081923.
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 5-EPI-ARISTOLOCHENE SYNTHASE.
GN EAS.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. BELL;
RA Zavala-Paramo G., Chavez-Moctezuma M.P., Garcia-Pineda E., Yin S.,
RA Chappell J., Lozoya-Gloria E.;
RT *Isolation of an elicitor-stimulated 5-epi-aristolochene synthase gene
RT (gpAS1) from chili pepper (Capsicum annuum).";
RL Physiol. plantarum 110:410-418(2000).
DR EMBL: AJ005588; CAA06614.1; -.
DR HSSP: Q40577; SEAT.
DR InterPro: IPR001906; Terp_synth_fam.
DR Pfam: PF01397; Terpene_synth; 1.
SQ SEQUENCE 559 AA; 65136 MW; EF866291AF12E88 CRC64;

Query Match 10.8%; Score 77; DB 10; Length 559;
Best Local Similarity 34.0%; Pred. No. 49;
Matches 18; Conservative 11; Mismatches 20; Indels 4; Gaps 2;

OY 73 YKFEFVLPDSAKIEVYDKNKKKEKSPITEKGFV---VPDSEHKN 122
DB 357 YKDYE-KEMSRDGRSHVYVYAKERLKLKLVKSYNIEAKWFIEGHMPASEYLKN 408

RESULT 21
099JZ1 PRELIMINARY; PRT; 608 AA.

AC 099JZ1.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROLACTIN RECEPTOR.
GN PRLR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strusberg R.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC005555; AAH05555.1; -.
DR EMBL: BC006652; AAH06652.1; -.
DR HSSP: P16471; 1BP3.

DR MGD; MG1:97763; PRLR.
DR InterPro: IPR002996; CRLA.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003528; Hematopo_receptor_L_F1.
DR Pfam: PF00041; fn3; 2.
DR SMART: SM00060; FN3; 2.
DR PROSITE: PS01352; HEMATOPO_REC_L_F1; UNKNOWN.1.
DR Receptor.
SQ SEQUENCE 608 AA; 68223 MW; 2710DAEC2B1A8F63 CRC64;

Query Match 10.8%; Score 77; DB 11; Length 608;
Best Local Similarity 23.9%; Pred. No. 54;
Matches 34; Conservative 17; Mismatches 43; Indels 48; Gaps 7;

OY 4 FDKGKYKGDASY-----FEPTGPY--LWVNTGVDSKGNELSP-HYVEP---I 49
DB 274 FDTLLERKGSSEILSALGODEPPTSDCDLVEFLVYDNDENRMLPSHSREYGGGV 333
OY 50 KPGTTLTKERIEYVEMALDATAKFEFVLPDSAKIEVYDKN---NKKKEETKSP 105
DB 334 KP-----THLPDSDSGSHSYSHSLSEKCEPOAVP 366

OY 106 ITEKGFVVPDSEHKNPGENL 127
DB 367 ---PAFHIPETKEPNPEANI 385

RESULT 22
029463 PRELIMINARY; PRT; 289 AA.

AC 029463.
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AE0795.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrleides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.R., Badger J.H., Glodex A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL: AE001049; AAB90443.1; -.
DR TIGR: AF0795; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 289 AA; 32544 MW; 12307FED8B25059C CRC64;

Query Match 10.7%; Score 76.5; DB 17; Length 289;
Best Local Similarity 27.9%; Pred. No. 25;
Matches 43; Conservative 15; Mismatches 59; Indels 37; Gaps 8;

OY 8 KYKGGDASYEPTG--PYLAV-NVTGVDSKGNELSPHY-----VEFPI 49
DB 32 YKKEGDRYVAKRPIGIGIPLAVNMITGVPTDGKSLABQRAITQAGSGYVLYVYESPA 91

OY 50 KPGTLTKETIEYVWALDATALYKERRVELDPSAKIEVITYDKNKKEETKSPITEK 109
 DB 92 NELYTAMERSE---AMGVDFTKV-ESNVVYVIDASESEDELK--BNPKALMETMAVATKEK 145
 OY 110 GFV-----VFDLSEH-----IKNCFENITK 130
 DB 146 KYNTIIDSITGLYEHKEMARQIROPFENMKK 179

RESULT 23

O971G3 PRELIMINARY: PRT: 349 AA.
 AC O971G3;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE SUN PROTEIN.
 GN ST1390.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 ON NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX PubMed-11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
 Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 Yoshikawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yanagishi A.,
 Ra Oshima T., Kikuchi H.;
 RT *Complete genome sequence of an aerobic thermacidophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";
 RL DNA Res. 8:123-140(2001).
 DR EMBL: AP000986; BAB66457.1; -
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 349 AA: 40344 MW: CAA5057BC19E815B CRC64;

Query Match 10.7%; Score 76.5; DB 17; Length 349;
 Best Local Similarity 25.5%; Pred. No. 31;
 Matches 36; Conservative 20; Mismatches 28; Indels 57; Gaps 9;

OY 34 SKGN---ELLSPHY-----VEPPIKGTLTKETIEYV-----EMALDATALY- 73
 DB 133 NKNLVVAVLNPXSAKFAADLEY-VKGSILQKASMYAHLLDPKPNKIIDMTATYP 191
 OY 74 --KERVVELDPSAKIEVITYDKNKKEETKSPITEKGFVVPDLSEHIKN----- 122
 DB 192 GSKLTHIYQLEPRSK--VIGFDHTRK-----VDELREKVKMKMNIETYL 235
 OY 123 -----PGFNL--ITKYVIE 134
 DB 236 ADSRLYEGFGLRDYDKYIID 256

RESULT 24

O9CFA0 PRELIMINARY: PRT: 252 AA.
 AC O9CFA0;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
 DE UNKNOWN PROTEIN.
 GN YOCB.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 ON NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IL403;
 RX MEDLINE-21235186; PubMed-11337471;

RA Bolotin A., Muncher P., Manger S., Jallion O., Malarne K.,
 RA Weissbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis ssp. lactis IL1403.";
 RL Genome Res. 11:731-753(2001).
 DR EMBL: AE006388; AAK05679.1; -
 KM Complete proteome.
 SQ SEQUENCE 252 AA: 27078 MW: BFDD38523950823B CRC64;

Query Match 10.6%; Score 76; DB 16; Length 252;
 Best Local Similarity 26.3%; Pred. No. 23;
 Matches 31; Conservative 12; Mismatches 29; Indels 46; Gaps 6;

OY 1 SSSFDKGRY-KGDDASYFEPTGPLYLVNVTGVDSKGNELLSPHYVEPPIKGTLTKETK 59
 DB 82 ASSFDENHKSINKDQTYLAKAKRYF-----DSEVLTNTYQVYDRGT----- 125
 OY 60 IEYVWALDATALYKERRVELDPSAKIEVITYDK-----NKKKEETKSF 104
 DB 126 ---FAGWNL-----KVE-----KTQVYQIDSGPQKVKLKGATINLKSETASF 167

RESULT 25

O91XY8 PRELIMINARY: PRT: 933 AA.
 AC O91XY8;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PROTOCOLADHERIN GAMMA A11.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE-99308636; PubMed-10380929;
 RA Wu Q., Mantatis T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes.";
 RL Cell 97:779-790(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE-20202599; PubMed-10716726;
 RA Wu Q., Mantatis T.;
 RT "Large exons encoding multiple ectodomains are a characteristic
 RT feature of protocadherin genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE-21154914; PubMed-11230163;
 RA Wu Q., Zhang T., Cheng J.-F., Kim Y., Grimwood J., Schmutz J.,
 RA Dickson M., Noonan J.P., Zhang M.Q., Myers R.M., Mantatis T.;
 RT "Comparative dna sequence analysis of mouse and human protocadherin
 RT gene clusters.";
 RL Genome Res. 11:389-404(2001).
 DR EMBL: AY013794; AAK26083.1; -
 SQ SEQUENCE 933 AA: 101124 MW: B6F0736CA125FD11 CRC64;

Query Match 10.5%; Score 75; DB 11; Length 933;
 Best Local Similarity 26.7%; Pred. No. 14e+02;
 Matches 36; Conservative 25; Mismatches 44; Indels 30; Gaps 9;

OY 20 PTGPLYLVNVTGV--DSKGNELLSPHYVEPPIKGTLTKETIEYVE-----MALD 69
 DB 466 PRGSSIF-SYTAIDRDSRENAQVITYSLAETIQ-GTPLSS-----YVINSIDPGVLYALQ 518
 OY 70 ATAYKERRVEL-----DP--SAKIEVITY--DKNKKKEETKSPITEKGFVVPDL 117

DB 519 SFEDYEFQTLQGLGVASDNGDPLSSNISLTFLVLDQNDNTPELLIPSLPTDGTGVELA 578
 QY 118 EHIKNPGFNLITRKV 132
 DB 579 PRSAEPGY-LVTRKV 592

RESULT 26
 P96018 PRELIMINARY; PRT; 666 AA.

AC P96018: 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DE MECA PROTEIN.
 GN MECA.
 OS Staphylococcus sciuri.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=1296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC29062;
 RA MEDLINE=97302569; PubMed=9158816;
 RX Wu S., Placitelli C., de Lencastre H., Tomasz A.;
 RT "Tracking the evolutionary origin of the methicillin resistance gene:
 RT a susceptible strain of Staphylococcus sciuri.";
 RL Microb. Drug Resist. 2:435-441(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC29062;
 RA Wu S.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y09223; CAAT0425.1; -
 DR InterPro: IPR001460; Transpeptidase.
 DR Pfam: PF00905; Transpeptidase; 1
 SQ SEQUENCE 666 AA; 75606 MW; CF046F9D10F29529 CRC64;

Query Match 10.4%; Score 74.5; DB 2; Length 666;
 Best Local Similarity 22.9%; Pred. No. 1e+02;
 Matches 33; Conservative 23; Mismatches 53; Indels 35; Gaps 6;

QY 27 VNVTVGDSKGNELSPHYVEFP-----IKPTTLTK-----EKI---EYVV---- 64
 DB 141 INEPLKSRGKILDRNNELATTTGTHVGIIVNNVSTDYKATAEKLDLSESYIKOOT 200
 QY 65 --EWALDAATVKEFRVVELDPSAKIEVTYYDKNKKKEFKSPITEK-----GFVVPDL 117
 DB 201 EODWVADDFVPLKTVQDNODKNVEYHLLTSQTESRQYPLEATTHLLGYGPIINS 260
 QY 118 EHIKNPGFN-----LITKVIETK 135
 DB 261 EELKQKAFKGYKKDAIVGKGIETK 284

RESULT 27
 O9XUQ2 PRELIMINARY; PRT; 233 AA.
 AC O9XUQ2: 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE T0865.9 PROTEIN.
 GN T0865.9.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Snye R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z81589; CAB04722.1; -
 SQ SEQUENCE 233 AA; 25912 MW; F40BCE0A2A1710E CRC64;

Query Match 10.3%; Score 74; DB 5; Length 233;
 Best Local Similarity 20.7%; Pred. No. 33;
 Matches 25; Conservative 28; Mismatches 50; Indels 18; Gaps 4;

QY 20 PTGPIYLVNNTGVDSKGNELSPHYVEFPKPTTLTKKIEYVENALATATYKRRV 79
 DB 91 PEPALRNNTVSLTSSDDIQOFFLE-----NESL-FOIEKYIDYKNRDTYIV 139
 QY 80 ----ELDPSAKIEVTYYDKNKKKEFKSPITEKGFVVPDLSEHIKNPGFNLITRKV 135
 DB 140 PSFGTVIPKSKCEIALICRPTGKSESQOMTISYK---TVDSNDNLANGCSEMEKIVIDI 196
 QY 136 K 136
 DB 197 K 197

RESULT 28
 O29007 PRELIMINARY; PRT; 233 AA.
 AC O29007: 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MEDLINE=98049343; PubMed=9389475;
 GN Archaeoglobus fulgidus.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 CC Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Richardson K.A., Dodson R.J., Gilm M., Hickey E.K., Peterson J.D.,
 RA Fieischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.J., McNeil L.K., Badger J.H., Glodak A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 DR EMBL: AE001017; AAB89982.1; -
 DR TIGR: AF1261; -
 DR InterPro: IPR000873; AMP-bind.
 DR Pfam: PF00501; AMP-binding; 2.
 KW Hypothetical protein; Ligase; Complete proteome.
 SQ SEQUENCE 233 AA; 27066 MW; 14C1ECAR62093FB6 CRC64;

Query Match 10.3%; Score 74; DB 17; Length 233;
 Best Local Similarity 22.7%; Pred. No. 33;
 Matches 32; Conservative 24; Mismatches 41; Indels 44; Gaps 7;

0Y 2 SSFDRKKKGGDASFEPPGYLMV--NVTGVDSKGNELLS-----41
 DB 97 NSFVDMKMSG-DVAIVITPEG-YKVVYDRVADIKSGGEMISVDLENYLGHVAVLEAC 154
 0Y 42 -----PHYVEPPI-----KGGTITIKRKI-----EYVENALDAAKKEKRYVELDSA 85
 DB 155 VVAHEHPKQEPRIIVVPPKPESEVTKDELREFLAKRPAAKQLPDD-----IIFVNEIP 208
 0Y 86 KIEVYYDNKKKKEETKSPFI 106
 DB 209 KTSVGRFDRKRLREQYRRYLI 229

RESULT 29
 09XU03 PRELIMINARY; PRT: 256 AA.

ID 09XU03:
 AC 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE TORGS.8 PROTEIN.
 GN TORGS.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
 OC Rhabditiida; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smye R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 investigating biology".
 RL Science 282:2012-2018(1998).
 DR EMBL; 281589; CAB04721.1; -;
 SQ SEQUENCE 256 AA; 28494 MW; 6BDFDFA05921D939 CRC64;

Query Match 10.3%; Score 74; DB 5; Length 256;
 Best Local Similarity 20.7%; Pred. No. 37;
 Matches 25; Conservative 28; Mismatches 50; Indels 18; Gaps 4;

0Y 20 PPGPYLMVNTGVDSKGNELSPHYVEPPIKGGTITIKRKIYYEMALDAAKKEKRYV 79
 DB 114 PEPALRSNVTSLTSESDDIQFLE-----NESL-FQIEIKIDIKNYDYIIV 162
 0Y 80 ----ELDPSAKIEVYYDNKKKKEETKSPITEKGFVVPDISEHIKNGFNLTITRVIEK 135
 DB 163 PPGFYVIPSKEICMILICRPGKSSDDQWTISYK---FVDSNDLANCGFSMEKIVIDI 219
 0Y 136 K 136
 DB 220 K 220

RESULT 30
 004622 PRELIMINARY; PRT: 462 AA.

ID 004622:
 AC 004622:
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE A_IG002N01.15 PROTEIN.
 GN A_IG002N01.15.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eudicots II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Schect P., Maggi L.;
 RT "The sequence of A. thaliana IG002N01.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Mash-U;
 RT "The A. thaliana Genome Sequencing Project.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Waterston R.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF007269; AAB61023.1; -;
 DR InterPro: IPR002213; UDPGT.
 DR Pfam: PF00201; UDPGT; 1.
 DR PROSITE: PS00375; UDPGT; 1.
 SQ SEQUENCE 462 AA; 50779 MW; F2D6BA0E70BF2DE7 CRC64;

Query Match 10.3%; Score 74; DB 10; Length 462;
 Best Local Similarity 28.6%; Pred. No. 74;
 Matches 38; Conservative 16; Mismatches 51; Indels 28; Gaps 7;

0Y 6 KGRYKGGDASFEPPGYLMVNTGVDSKGNELLS-----LISPHYVEPPIKPGF 53
 DB 206 KALDEPGIDRPVYVGP--LVNIGKQEAQTESSECLKMLDQPLGSLVYVF--GSGG 261
 0Y 54 TLTKKIEYYEMALD-ATAYKKEKRYVELDPSAKIEVYYDNKKKKEETKSPITEKGFV 112
 DB 262 TLTCQQLN---ELALGLADSEQRFLWYIRSPSGIANSYFDSHSQD-----FLT---FL 310
 0Y 113 VPDISEHIKNGF 125
 DB 311 PPGFLERTKRGF 323

RESULT 31
 09M156 PRELIMINARY; PRT: 480 AA.
 ID 09M156:
 AC 09M156:
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE FLAVONOL GLUCOSYLTRANSFERASE.
 GN ATG01070.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eudicots II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA ED Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.W., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamita A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narasaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full Length cDNA of gene ATG01070 (GI:7267604).";

RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
 RA Quesh H.L., Tang C., Toriumi M., Yu G., Bowser L., Carrinot P.,
 RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamlya A.,
 RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT *Full length cDNA of gene At4g01070 (GI:7267604).
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL161491; CAB80916.1; -
 DR EMBL: AF360262; AAK25972.1; -
 DR EMBL: AY040075; AAK64133.1; -
 DR Interpro: IPR002213; UDPGT.
 DR PROSITE: PS00375; UDPGT; 1.
 KW Transferase.
 SQ SEQUENCE 480 AA; 52923 MM; 3E0315C1D71D2DB0 CRC64;

Query Match 10.3%; Score 74; DB 10; Length 480;
 Best Local Similarity 28.6%; Pred. No. 78;
 Matches 38; Conservative 16; Mismatches 51; Indels 28; Gaps 7;

OY 6 KCKYKGGDASFEPTGYPLMVNTGVSCKNE-----LLSPHYVEPIKPGT 53
 DB 224 KALQEGDLKPPYVGP--LVNIGKQKAEQTESECLKWLQNLGSLVYVSF--GSGG 279
 OY 54 TLTKKEIYEVNADL-ATAYKREFRYVELDPSAKIEVYTDKKNKEETSPITEKGFV 112
 DB 280 TLTKEDLN---ELALDLADSEORFLWVNSPISANSYFSDSHSQD---PLT---FL 328
 OY 113 VPDLSHINKPGF 125
 DB 329 PEGFLERTKRGF 341

RESULT 32
 O9PPW0 PRELIMINARY; PRT; 552 AA.
 AC O9PPW0:
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE PHOSPHOMANNOMUTASE.
 GN CP5G OR U0530.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Ureaplasma.
 OC NCBI_TaxID=134821;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN-SEROVAR 3:
 RX MEDLINE=20500219; PubMed=11048724;
 RA Class J.I., Letkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
 RA Caselli G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RT urealyticum."
 RL Nature 407:757-762(2000).
 DR EMBL: AE002152; AAF30943.1; -
 DR InterPro: IPR001485; PGM_PMM.
 DR Pfam: PF02878; PGM_PMM_1; 1.
 DR Pfam: PF02879; PGM_PMM_1; 1.
 DR PRINTS: PR00509; PGM_PMM.
 DR PROSITE: PS00710; PGM_PMM; 1.
 KW Complete proteome.
 SQ SEQUENCE 552 AA; 64308 MM; F3FB5E15EC98DE0 CRC64;

Query Match 10.3%; Score 74; DB 16; Length 552;
 Best Local Similarity 26.5%; Pred. No. 92;

Matches 22; Conservative 17; Mismatches 32; Indels 12; Gaps 3;
 OY 47 FPIKPGTTLTKREIY-----YEWALDATAYKEFRVELDPSAKIEVYTD--KNKKE 99
 DB 471 FEKRTVQYVTHNIHIEFGSLEYEMKLEYSSTIKFRLSGTEPKPKVYIDLYDSKQKRD 530
 OY 100 ETKSFPTTEKGFVVPDISEHIKN 122
 DB 531 FYQELKIKAE-----DILDFLKN 548

RESULT 33
 O92N17 PRELIMINARY; PRT; 257 AA.
 AC O92N17:
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE HYPOTHETICAL PROTEIN SMC01600.
 GN SMC01600.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=1021;
 RX MEDLINE=21368234; PubMed=11474104;
 RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
 RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
 RA Boutry M., Bowser L., Buhrmester U., Cadieu E., Capela D., Chailin P.,
 RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
 RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gujral M.,
 RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
 RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
 RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetalle D., Purnelle B.,
 RA Rampeger U., Surzycki R., Thebaud P., Vandenbol M.,
 RA Vorholter F.J., Weidner S., Wells D.H., Wong K., Yen K.-C., Batut J.;
 RT "The composite genome of the legume symbiont Sinorhizobium meliloti."
 RL Science 293:668-672(2001).
 DR EMBL: AL591790; CAC46794.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 257 AA; 27978 MM; B9488C3C411FAD2E CRC64;

Query Match 10.3%; Score 73.5; DB 16; Length 257;
 Best Local Similarity 23.6%; Pred. No. 41;
 Matches 30; Conservative 18; Mismatches 62; Indels 17; Gaps 4;
 OY 27 VNTGTGDSKGNELSPHYV--EPIKPGT-LTKREIYEVNADLATAKREFRYELDP 83
 DB 13 VAENMAQKGEFLVCEHASNRMPESLGTGLSKALESHVAMPALAVSRLLCEKLDG 72
 OY 84 S---AKIEVYTDKKNKEETKSPITEKGFV-----DSEHIKNPGFNLT 129
 DB 73 TLHQRSLATYDCNRPEDDAMPVLSYVDYVGNMSKSAERKSRVDEILYLFHNAV 132
 OY 130 KVYIEKK 136
 DB 133 RAVADRK 139

RESULT 34
 O9XIR2 PRELIMINARY; PRT; 431 AA.
 AC O9XIR2:
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE FL3011.13 PROTEIN.
 GN FL3011.13.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Federfpiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altafi H., Araujo R., Hultzer L., Rowley D., Buehler E., Dunn P.,
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
 RA Lucero S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.,
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
 DR EMBL: AC006193; AAD38257.1; -
 DR HSSP: P00799; 2ASI.
 DR MEROPS: A01.0PW; -
 DR InterPro: IPR002328; ADH_zinc.
 DR InterPro: IPR001969; Asp_protease.
 DR InterPro: IPR001461; Pepsin.
 DR Pfam: PF00026; asp_1.
 DR PRINTS: PR00792; PEPsin.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_2.
 KW Oxidoreductase; Zinc.
 SQ SEQUENCE 431 AA; 46340 MW; 644FF17BA386ECA CRC64;

Query Match 10.3%; Score 73.5; DB 10; Length 431;
 Best Local Similarity 26.3%; Pred. No. 76;
 Matches 35; Conservative 17; Mismatches 52; Indels 29; Gaps 6;

QY 11 KGDASYFEPTGYLMVNTGVDSKGNELSPHY---VEFPKPGTTLTKKIEYVW 66
 DB 275 KKDPAITY-----PLNEAISVSGSKIQFTSTFGEGNIVDSGTLTLPSNFYEL 329
 QY 67 -ALDANKERRYVELDPSAKIEVITYYDKNKKRETSFPTTEGFPVPLSEIKKPGF 125
 DB 330 ESYVAATKARV--QDPDGLSLCYRDS-----SFPVPLTIVFKGGDV 373

QY 126 ---NLTKVIEK 135
 DB 374 KLGNIINFVANSE 386

RESULT 35
 042976 PRELIMINARY; PRT; 764 AA.
 AC 042976;
 DT 01-JAN-1999 (TREMBlrel. 09, Created)
 DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHEICAL. 85.4 KDA PROTEIN C20F10.07 IN CHROMOSOME II.
 GN SPC20F10.07.
 OS Schizosaccharomyces pombe (Fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Lyne M., McDougall R., Rajandream M.A., Barrell B.G., Beck A.,
 RA Reinhardt R.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: SOME, TO YEAST YHR080C.
 DR EMBL: AL021747; CA16847.1; -
 DR InterPro: IPR004182; GRAM.
 DR Pfam: PF02893; GRAM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 764 AA; 85354 MW; 080990ESC7FA7B2 CRC64;

Query Match 10.3%; Score 73.5; DB 3; Length 764;

Best Local Similarity 23.0%; Pred. No. 1.5e+02;
 Matches 34; Conservative 16; Mismatches 49; Indels 49; Gaps 6;

QY 7 GKXKGGD-----ASYEPTGP-----YLMVNTGVDSKGN-ELLS-----PHYE 46
 DB 475 GKWEKIDDKMKNRKYIKPAPYROTSCYITDTIQHLDINNIEILSTTPDVPSGTS 534
 QY 47 FPKI-----PGTTLTKKIEYVEMALD-----TAVKEER 77
 DB 535 FVVKITLYALSMHSSKTKLNISYVEMSKSNKGPTEKAGQAGQASYKDLTFENRK 594

QY 78 VVELDPSAKIEVITYYDKNKKRETSF 105
 DB 595 VSPKGRKRKTKTKTKKKKNHASETSVAP 622

RESULT 36
 091X25
 ID 091X25 PRELIMINARY; PRT; 792 AA.
 AC 091X25;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PROTOCADHERIN BETA 5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RC MEDLINE=99308636; PubMed=10380929;
 RA Wu Q., Maniatis T.;
 RT "A striking organization of a large family of human neural cadherin-
 RL like cell adhesion genes.";
 RL Cell 97:779-790(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RC MEDLINE=20202599; PubMed=10716726;
 RA Wu Q., Maniatis T.;
 RT "Large exons encoding multiple ectodomains are a characteristic
 RL feature of protocadherin genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RC MEDLINE=21154914; PubMed=11230163;
 RA Wu Q., Zhang T., Cheng J.-F., Kim Y., Grimwood J., Schmutz J.,
 RA Dickson M., Noonan J.P., Zhang M.Q., Myers R.M., Maniatis T.;
 RT "Comparative dna sequence analysis of mouse and human protocadherin
 RL gene clusters.";
 RL Genome Res. 11:389-404(2001).
 DR EMBL: AY013787; AAK26076.1; -
 SQ SEQUENCE 792 AA; 87272 MW; F7A313143227BB74 CRC64;

Query Match 10.3%; Score 73.5; DB 11; Length 792;
 Best Local Similarity 25.3%; Pred. No. 1.6e+02;
 Matches 39; Conservative 17; Mismatches 55; Indels 43; Gaps 8;

QY 13 DDAAYEPTGYLM-----VNVTVGVDKGN-----ELLSPHYEFPKPGT 53
 DB 445 DNAPAFSGTS-YTMFRENNSPALHIGTISATDSGNSNAHIYSLPPHDOALHSLI 503
 QY 54 TLTKREKIEYVEMALDPAVKEFRV-----ELDPSAKIEVITYYDKNKKKE----- 99
 DB 504 SINADNGOLFALSDLYEALQGFYVVGATDRGSPELSSQALVRYVVDNDNAPFLXP 563

QY 100 -ETKSPITEKGFVPVPLSEIKKPGNLTIKYV 132
 DB 564 LQNNAPCTE---VLPRAAE---PGI-LVTKVY 589

```

RESULT 37
ID 09X079 PRELIMINARY; PRT; 1536 AA.
AC 09X079;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE HYDROTHERMAL 176.0 KDA PROTEIN.
GN TM0984.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Half D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Ulterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 393:323-329(1999).
DR EMBL: AE001760; AAD36063.1; .
DR TIGR: TM0984; .
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 1536 AA; 175955 MW; 71DA93908F0CC1F4 CRC64;

Query Match
Best Local Similarity 10.3%; Score 73.5; DB 16; Length 1536;
Matches 36; Conservative 13; Mismatches 44; Indels 57; Gaps 6;

QY 3 SPDKGKKKDDASVFEPTGYLMVNTGVDSKGNELSPHYE-----VEPPIP 51
DB 234 SFEGLYKA-----FEKT-----KVTVSIFPTKKNEVSEETDELGFSGSKMLPDRA 283
QY 52 GTTLTK-----EKIEYVEMALDATAYKEFVVEL-----DP 83
DB 284 SVGLKVVNDHGRARYEFLVEYRKPKPEYKIEITDKVYISGEVYVNLVRYKNGCP 343
QY 84 SAKIEYTYDDKKKKKEKTSPTTEKGVV 113
DB 344 VAKAQVAY-----VRAPEEGSGYLV 365
GN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

```

```

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombaris P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch R.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK015254; BAB29766.1; .
DR MGD: MGI:1921888; 4930430K04Rik.
DR InterPro: IPR000313; PWM.
DR Pfam: PF00855; PWM.
DR SMART: SM00293; PWM.
FT NON-UTR 1
SQ SEQUENCE 171 AA; 19652 MW; 4E9F0BB5BF60B7A6 CRC64;

```

```

Query Match
Best Local Similarity 10.2%; Score 73; DB 11; Length 171;
Matches 35; Conservative 15; Mismatches 38; Indels 52; Gaps 7;

QY 6 KGRKRRKDDASVFEPTGYLMVNTGVDSKGNELSPHYEPPKPGT----- 54
DB 53 KGRY-----VT-YDQGN--VEKYVEFLGDHSHRMSAARVGHF 90
QY 55 -LTKIEYVEMALDATAYKE-----FVVELDSAKIEYTYDK-NKKKEETS 103
DB 91 STLE-----AADCTKKRWYRSALBEAYOLYRCASAEORLEVCCLSPKPAKDTYA 142
QY 104 FYTEKGVVVDSEHIKNP 123
DB 143 AVYTKRGMVSKINTEKKRP 162

```

```

RESULT 39
ID 09I946 PRELIMINARY; PRT; 345 AA.
AC 09I946;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SPINAL CORD-DERIVED GROWTH FACTOR.
GN SCDGF.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN; TISSUE=SPINAL CORD;
RX MEDLINE=20317014; PubMed=10858496;
RA Hamada T., Ue T., Miyata Y.;
RT "A novel gene derived from developing spinal cords, SCDGF, is a unique
RT member of the PDGF/VEGF family."
RL FEBS Lett. 475:97-102(2000).
CC 1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: AB033829; BAB03265.1; .
DR InterPro: IPR000859; CUB.
DR InterPro: IPR000072; PDGF.
DR Pfam: PF00431; CUB.
DR SMART: SM00042; CUB.
DR SMART: SM00141; PDGF.
DR PROSITE: PS01180; CUB.
DR PROSITE: PS00278; PDGF_2.
SQ SEQUENCE 345 AA; 38940 MW; 97ACEA992BF5128C CRC64;

```

```

Query Match      10.2%; Score 73; DB 13; Length 345;
Best Local Similarity 24.7%; Pred. No. 65;
Matches 40; Conservative 17; Mismatches 41; Indels 64; Gaps 9;

OY 1 SSSFDKGGKYGKGDAA-----SYEPTGPIYLVNTGVDSKGNELLSPHYVEFPKIR--- 50
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 126 SSSVPSRQISKGNQIRIRFVSDEYF-HSQPFCIHYY-----LVPHTTEAPSSSL 176
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
OY 51 PGTTL-----TKKIEYYE-----WALDATAY---KEFRVV 79
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 177 PPSALPLDVNNNAVAGSTVELIRYLEPDRMQDLEDLYRPTQGLGKAVIHGRKSRV 236
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
OY 80 ELDPASAKLEVITYRNKKKTEKSPITEKGFVVPDISEHK 121
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 237 DL-----NLKKEVRLYSCTPRNFVS-SLREELK 264
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 40
OY 0973T1 PRELIMINARY; PRT; 419 AA.
AC 0973T1;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE PUTATIVE GLUCOSE-1-PHOSPHATE NUCLEOTIDYLTRANSFERASE.
GN ST0816.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT *Complete genome sequence of an aerobic thermacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.*;
RL DNA Res. 8:123-140(2001).
RW EMBL; AP000963; BAB65829.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 419 AA; 47853 MW; 8F05B03DD674DD73 CRC64;

Query Match      10.2%; Score 73; DB 17; Length 419;
Best Local Similarity 27.0%; Pred. No. 82;
Matches 30; Conservative 24; Mismatches 35; Indels 22; Gaps 7;

OY 17 YEETPGPIYLVNTGVDSKGNELLSPHYVEFPKIRGTTLTKKIEYYEMALDATAY 73
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 112 YVDKPEPLVQGNLFYLDVK--KVEFYHEEK---KAFMTIVLAKVE-----SV 156
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
OY 74 KEFRVELDPSAKLEVITYRNKKKTEKSPITEKGFV---PDLSEHIKN 122
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 157 EEFGVAELNDMKIK-REVEKPRREAP-S-DLANTGIYISPEIRDYFKS 205
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 41
OY 0227J2 PRELIMINARY; PRT; 470 AA.
AC 0227J2;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE F11P17.14 PROTEIN.
GN F11P17.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.

```

```

OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RA Federespelet N.A., Conway A.B., Conway A.R., Davis K., Palm C.J., Au M.,
RA Araujo R., Chung E., Kutz D.B., Buehler E., Dewar K., Feng J.,
RA Kim C., Li Y., Shinn P., Sun H., Oji O., Osborne B., Shen Y.K.,
RA Toriumi M., Vytaskeala V., Yu G., Theologis A., Ecker J., Davys R.W.;
RT *Genomic sequence of Arabidopsis BAC F11P17.*;
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC002294; AAB71481.1; -.
DR HSSP; P08819; 1MHT.
DR InterPro; IPR000379; Est_1lp-thioest_actsite.
DR InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF00450; serine_carbpept; 2.
DR PRINTS; PR00724; CRBOXYPRASEC.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 470 AA; 53687 MW; 2806E3C3EE7116B2 CRC64;

Query Match      10.2%; Score 73; DB 10; Length 470;
Best Local Similarity 26.8%; Pred. No. 94;
Matches 33; Conservative 14; Mismatches 38; Indels 38; Gaps 7;

OY 22 GPIYLVNTGVDSKGNELLSPHYVEFPKIRGTTLTKKIEY----- 62
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 98 GPPL-----VDNKGNSLKNFPYAMNKEANILFLESFAGVGSYSNTSSDYRLGDPTA 151
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
OY 63 -----YEWALDATAYKE--FRVELDPSAK-----IEVYYDNKKKTEKSPITEKG 110
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 152 RDSYTFLOKKNFLFPAYKEDFPIAGEYAGKYVPELAEYIDKNDENL-SLHINLKG 210
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
OY 111 FVV 113
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 211 ILV 213

RESULT 42
OY 09FRF0 PRELIMINARY; PRT; 654 AA.
AC 09FRF0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE HYPOTHETICAL 73.5 KDA PROTEIN.
GN OSJNB0009F04.15.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
RA Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Craven B., Uteback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT *Oryza sativa chromosome 3 BAC OSJNB0009F04 genomic sequence.*;
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC079630; AAG46073.1; -.
KW Hypothetical protein.
SQ SEQUENCE 654 AA; 73476 MW; C758A457048E74BA CRC64;

Query Match      10.2%; Score 73; DB 10; Length 654;
Best Local Similarity 23.2%; Pred. No. 14e+02;
Matches 32; Conservative 21; Mismatches 41; Indels 44; Gaps 7;

OY 1 SSSFDKGGKYGKGDASYEPTGPIYLVNTGVDSKGNELLSPHYVEFPKIRGTTLTKK 59
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 431 SASSSSTFFVKGPPA---PPAHTATKTKAKDAKNK-----DPYDCTDGE 476
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
OY 60 IEVYEMALDATAYKE-----RVVELDPSAK-----IEVYYDNKKKKEE 100
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

DB 477 LDYVAV-SEVKQKPKSPKPKIPIDSVNPFNGMSASVKAIKLSYERTLTK-- 530
QY 101 TKSPITEKGFVVDLSE 118
DB 531 ----ATSGSKTYVPLQGE 544

RESULT 43
Q9H706 PRELIMINARY: PRT: 875 AA.
ID 09H706:
AC 09H706:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA: FLJ21610 FIS, CLONE COL07333.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki T., Odayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isoigai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK025263; BABI5094.1;
DR InterPro: IPR001660; SAM.
DR SMART: SM00454; SAM; 1.
SQ SEQUENCE 875 AA; 97099 MW; 10A0E9147A9BFCB1 CRC64;

Query Match 10.2%; Score 73; DB 4; Length 875;
Best Local Similarity 30.4%; Pred. No. 26+02;
Matches 24; Conservative 15; Mismatches 20; Indels 20; Gaps 4;

QY 43 HYVEPIKPGTTLTKKEIYVEMALDPAVKEFVVELDPSAKIYVYDKKKKEETK 102
DB 73 HYVIGP-----KIEIPVHYA-----GQFKLEDDRDRIKEVQYF--NSVEVAK 114

QY 103 SPITEKGFVVDLSEHK 121
DB 115 AFP--ERYVMEDITFNK 131

RESULT 44
Q63577 PRELIMINARY: PRT: 2505 AA.
ID 063577:
AC 063577:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FATTY ACID SYNTHASE.
GN FATTY ACID SYNTHASE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE-89240666; PubMed-2717611;
RA Amy C.M., Wilkowski A., Naggert J., Williams B., Randhawa Z.I.,
RA Smith S.;
RT "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT fatty acid synthase."
RL Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE-92141210; PubMed-1736293;

```

```

RA Smith S., Naggert J.K., Williams Ahlf B., Amy C.M.;
RT "Intion-exon organization of the gene for the multifunctional animal
RT fatty acid synthase."
RL Proc. Natl. Acad. Sci. U.S.A. 89:1105-1108(1992).
DR EMBL: M84761; AAA41145.1;
DR InterPro: IPR001227; Acyltransf_domain.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; Phosphopant_attach.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl_transf. 1.
DR Pfam: PF00107; adh_zinc. 1.
DR Pfam: PF00109; ketoacyl-synt. 1.
DR Pfam: PF02801; ketoacyl-synt_C. 1.
DR Pfam: PF00550; pp-binding. 1.
DR Pfam: PF00975; Thioesterase. 1.
DR PROSITE: PS50075; ACP_DOMAIN. 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE. 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE. 1.
DR Phosphopantetheine; Transferase.
SQ SEQUENCE 2505 AA; 272663 MW; 4F97CFA5E83EEB78 CRC64;

Query Match 10.2%; Score 73; DB 11; Length 2505;
Best Local Similarity 30.0%; Pred. No. 6.9e+02;
Matches 24; Conservative 14; Mismatches 32; Indels 10; Gaps 4;

QY 27 VNVTVDSKGNELSPHVEPIKPGTTLTKKEIYVEMALDPAVKEFVVELDPS-S 84
DB 803 VHLTGIDINPNALPP--VEFVPRGTPPL-----ISPHIKW--DHSQWIDJPVADPFNGS 854

QY 85 AKIEVYTYDKKKKEETKSF 104
DB 855 SSSATVYVNIIDASSSSDHY 874

RESULT 45
Q16702 PRELIMINARY: PRT: 2509 AA.
ID 016702:
AC 016702:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FATTY ACID SYNTHASE (EC 2.3.1.85).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Hennigar R.A., Jenner K.H., Helne H.S., Kayler A.E., Wood F.D.,
RA Kunaide F.P., Pasternack G.R.;
RT "Molecular Cloning of Tumor-Associated Human Fatty Acid Synthase."
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U29344; AAA73576.1;
DR InterPro: IPR001227; Acyltransf_domain.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; Phosphopant_attach.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl_transf. 1.
DR Pfam: PF00107; adh_zinc. 1.
DR Pfam: PF00109; ketoacyl-synt. 1.
DR Pfam: PF02801; ketoacyl-synt_C. 1.
DR Pfam: PF00550; pp-binding. 1.
DR Pfam: PF00975; Thioesterase. 1.
DR PROSITE: PS50075; ACP_DOMAIN. 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE. 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE. 1.
DR Acyltransferase; Phosphopantetheine; Transferase.
SQ SEQUENCE 2509 AA; 273089 MW; 689018A8219C13B7 CRC64;

```

Query Match 10.2%; Score 73; DB 4; Length 2509;
Best Local Similarity 34.4%; Pred. No. 6.9e+02;
Matches 21; Conservative 10; Mismatches 28; Indels 2; Gaps 1;

QY 27 VNVGVDSKGNELSPHYVEFPPIKPTTLTKETKEIYYEMALDPAKKEFVELDSAK 86
DB 803 LHLSDIDANPALPPP--VEFPAPRGTPPLISPLIKMDSLAMDVPAADFPNGSGSPSA 860
OY 87 I 87
DB 861 I 861

RESULT 46
O9V4P4 PRELIMINARY; PRT; 2748 AA.

ID O9V4P4
AC O9V4P4; O9V4P3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE BCDA:GH02712 OR CG3088 OR CG17983.
GN BCDA:GH02712 OR CG3088 OR CG17983.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
Deery Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Fertlera S., Fleischmann W.,
Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jajani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacלב J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradlin A.C., Stapleton M., Strung R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan K.C., Zhu D., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
The genome sequence of Drosophila melanogaster.;
RT Science 287:2185-2195(2000).
DR EMBL: AE003840; AAF59222.2; -
DR FlyBase: FBgn0027604; BCDA:GH02712.

SQ SEQUENCE 2748 AA; 296242 MW; 922846A92CAABF9C CRC64;

Query Match 10.2%; Score 73; DB 5; Length 2748;
Best Local Similarity 28.4%; Pred. No. 7.7e+02;
Matches 31; Conservative 22; Mismatches 40; Indels 16; Gaps 5;

QY 20 PTGPTLVNVTGVDSKGNELSPHYVEFPPIKPTTLTKETKEIYYEMALDPAKKEFV 79
DB 1260 PEKPEPEMENTEIEE--EKVSPTSVE--IVPTTDLQASVETAPVSAVEEVPKI 1315
OY 80 ELDSAKIEVT-----YYDKNKKKEETKSPITKEKGVVVDLSEHI 120
DB 1316 E-DSTAIDKSPQISPIEGEHSQHSSEKTPES--VEVGIDSPNVADVI 1360

RESULT 47
Q29621 PRELIMINARY; PRT; 4968 AA.

ID Q29621
AC Q29621;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARDIAC RYANODINE RECEPTOR.
GN RYR-2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
[1]
SEQUENCE FROM N.A.
RC TISSUE=HEART, AND BRAIN;
RX MEDLINE=91032060; PubMed=2226801;
RA Nakai J., Imagawa T., Hakamat Y., Shigekawa M., Takeshima H., Numa S.;
RT "Primary structure and functional expression from cDNA of the cardiac
RYR2 receptor/calcium release channel.";
RL FEBS Lett. 271:159-177(1990).
[2]
SEQUENCE FROM N.A.
RC TISSUE=HEART, AND BRAIN;
RX MEDLINE=93147068; PubMed=1337083;
RA Imagawa T., Nakai J., Takeshima H., Nakasaki Y., Shigekawa M.;
RT "Expression of Ca(2+)-induced Ca2+ release channel activity from
cardiac ryanodine receptor cDNA in Chinese hamster ovary cells.";
RL J. Biochem. 112:508-513(1992).
[3]
SEQUENCE FROM N.A.
RC TISSUE=HEART, AND BRAIN;
RA Nakai J.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -/- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
DR EMBL: U50465; AAA93465.1; -
DR EMBL: U50465; AAA93465.1; -
DR InterPro: IPR000636; Cation_chan_non_119.
DR InterPro: IPR001682; Channel_pore_Ca_Na.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003608; MIR.
DR InterPro: IPR001215; Ryanodn_receptor.
DR InterPro: IPR000699; RYDR_IPPR.
DR InterPro: IPR003033; RYR.
DR InterPro: IPR003877; SPRY.
DR InterPro: IPR003878; SPRY_domain.
DR Pfam: PF00036; efhand. 2.
DR Pfam: PF00520; ion_trans. 1.
DR Pfam: PF02815; MIR. 4.
DR Pfam: PF01365; RYDR_IPPR. 2.
DR Pfam: PF02026; RYR. 4.
DR Pfam: PF00622; SPRY. 3.
DR PRINTS: PR00795; RYANDINER.
DR SMART: SM00472; MIR. 4.
DR SMART: SM00449; SPRY. 3.
KW Calcium-binding
SQ SEQUENCE 4968 AA; 565081 MW; 04B16A35D06B852D CRC64;

Query Match 10.2%; Score 73; DB 6; Length 4968;
 Best Local Similarity 25.0%; Pred. No. 1.6e+03;
 Matches 27; Conservative 20; Mismatches 35; Indels 26; Gaps 6;

OY 17 YEPTGCHLVNVTGCDKGNELSPHYVEPIKPTTLTKKEIEYV-----EWA 67
 DB 2680 YMSNVYMSMEKQSDSESN--FNOQ---PVDTSNIIPKELEFYINKYAEHSHDKWS 2733
 OY 68 LDFTA---YKEFRVVELDPSAKIE--VYYDKNKKKEETKSEPTTE 108
 DB 2734 MDKLANWITGEL-----YSDSKIQPLMKPKLSEKEKEIYWPPIKE 2777

RESULT 48
 O9XG34 PRELIMINARY; PRT; 237 AA.

AC O9XG34; 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 26S PROTEASOME IOTA SU.
 GN PRSAI.
 OS Gullardia theta (Cryptomonas phl).
 OC Eukaryota; Cryptophyta; Cryptomonadaceae; Gullardia.
 OX NCBI_TaxID=55529;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20087226; PubMed=10618395;
 RA Zauner S., Fraunholz M., Wastl J., Penny S.L., Beaton M.,
 Cavalier-Smith T., Mäler U., Douglas S.;
 RT "Chloroplast protein and centrosomal genes, a tRNA intron, and odd
 telomeres in an unusually compact eukaryotic genome, the cryptomonad
 nucleomorph.";
 RT Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21223349; PubMed=11323671;
 RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.,
 Wu X., Reith M., Cavalier-Smith T., Mäler U.;
 RT "The highly reduced genome of an enslaved algal nucleus.";
 RL Nature 410:1091-1096(2001).
 DR EMBL: AJ010592; CAB40400.1; -
 DR MEROPS: T01.971; -
 DR InterPro: IPR001353; Proteasome.
 DR Pfam: PF00227; proteasome; 1.
 KW Proteasome.
 SQ SEQUENCE 237 AA; 27567 MW; 2B331725BAD2580 CRC64;

Query Match 10.1%; Score 72.5; DB 10; Length 237;
 Best Local Similarity 28.6%; Pred. No. 46;
 Matches 28; Conservative 19; Mismatches 34; Indels 17; Gaps 4;

OY 7 GKRRKGDASYEPTGCHLVNVTGCDKGNELSPHYVEPIKPTTLTKKEIEYVEM 66
 DB 139 GQEFKGN-----TGSFILTNCISIGYKDI-----IEYLKKRKPESFONIEKSVK 186
 OY 67 ALDA--TAYKEFRVVELDPSAKIEVYVYDKNKKKEETK 102
 DB 187 LLEVLTISINDELVIE---NIRLFTSGDKNNVKKMESE 221

RESULT 49
 O976M4 PRELIMINARY; PRT; 311 AA.

AC O976M4; 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE NAD-DEPENDENT ALCOHOL DEHYDROGENASE.
 GN ST0075.

OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 Sekine M., Baba S.-I., Ankal A., Kosugi H., Hosoyama A., Fukui S.,
 Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";
 RL DNA Res. 8:123-140(2001).
 DR EMBL: AP000981; BAB55032.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 311 AA; 34546 MW; 883FC51752964D2A CRC64;

Query Match 10.1%; Score 72.5; DB 17; Length 311;
 Best Local Similarity 28.8%; Pred. No. 64;
 Matches 34; Conservative 14; Mismatches 39; Indels 31; Gaps 6;

OY 37 NELLSPHYVEPIKPGT---LTKEIEYVEMALD--ATAYKEFRVVELDPSAKIEVT- 90
 DB 99 NQNTSGGYAAEEYIVPETMLIPLPDEKFEKYAA-ALDPVATAIHATKVELNDESKVLVTG 157
 OY 91 -----YDANKKKEETKSPITRKGVVDLSHINKPQNL 128
 DB 158 GGGGVGHLIOYLYKLVGEVYALTSKIDKVEF--TE--YIINDVKGRPDVVEFLV 211

RESULT 50
 O674Z9 PRELIMINARY; PRT; 584 AA.

AC O674Z9; 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE TRANSPORTER (EXTRACELLULAR SOLUTE BINDING PROTEIN FAMILY 5).
 GN OPR4 OR AQ.1441.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulaj M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RL aeolicus.";
 RL Nature 392:353-358(1998).
 DR EMBL: AE000740; AAC07389.1; -
 DR InterPro: IPR001091; N4_Mtase.
 DR Pfam: PF00496; SBP_bac_5; 2.
 DR PROSITE: PS00093; N4_MTASE; UNKNOWN_1.
 DR PROSITE: PS01040; SBP_BACTERIAL_5; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 584 AA; 67509 MW; D9ED260E32D0C726 CRC64;

Query Match 10.1%; Score 72.5; DB 16; Length 584;
 Best Local Similarity 31.6%; Pred. No. 1.4e+02;
 Matches 36; Conservative 10; Mismatches 27; Indels 41; Gaps 8;

OY 19 EPTGYLVNVTGVDKGNELSPHYVEPIKPGT-----TLTKR-----KIEY 63
 DB 179 EPPAFPL-----NSLGEVILPKHKLKNYKEGTFSTAMVNTNPKELVGTGAYKIEY 231

QY 64 VEMALDATAKKEPRVVEIDPSAKIEVTYYDKNNKKEETKSEP-ITEK-GFVVPD 115
Db 232 V-----KGQRVV-----YRANPYYYEED---ENGNRRLPYIKKEKGVIIIPD 268

Search completed: September 1, 2002, 03:27:09
Job time: 562 sec

•
•
•
•

seq_name: gb_ba:AF003135	seq_documentation_block:	291150 bp	DNA	linear	BCT 07-FEB-2002
LOCUS	AP003135				
DEFINITION	Staphylococcus aureus subsp. aureus N315 genomic DNA, complete genome, section 7/10.				
ACCESSION	AP003135	BA000018			
VERSION	AP003135.2	GI:14349227			
KEYWORDS					
SOURCE					
ORGANISM	Staphylococcus aureus subsp. aureus N315 (sub-species:aureus N315, strain:N315) DNA.				
REFERENCE	Staphylococcus aureus subsp. aureus N315 Bacteria; Firmicutes; Bacilli/Clostridium group; Bacillus/Staphylococcus group; Staphylococcus.				
AUTHORS	1 (sites)				
TITLE	Kuroda,M., Ohta,T., Uchiyama,T., Baba,T., Yuzawa,H., Kobayashi,T., Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Iida,J., Ito,T., Kanamori,M., Matsunaru,H., Maruyama,A., Murakami,H., Hosoeyama,A., Mitsuani-Oi,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C., Sekimizu,K., Hiraoka,H., Kohara,S., Goto,S., Yabuzaki,J., Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C., Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiratazu,K.				
JOURNAL	Whole genome sequencing of methicillin-resistant Staphylococcus aureus				
MEDLINE	Lancet 357 (9264), 1225-1240 (2001)				
REFERENCE	2131952				
AUTHORS	2 (bases 1 to 291150)				
TITLE	Aoki,K., Oguchi,A., Hosoeyama,A., Nagai,Y., Kuroda,M., Hiratazu,K. and Kikuchi,H.				
JOURNAL	Direct Submission				
COMMENT	Submitted (30-JAN-2001) Kenichi Aoki, National Institute of Technology and Evaluation, Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan				
FEATURES	(E-mail:aoki-kenichi2@niet.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-1972, Fax:81-3-3481-8424)				
SOURCE	On Jun 12, 2001 this sequence version replaced gi:13701545. Location/Qualifiers				
	1..291150				
	/organism="Staphylococcus aureus subsp. aureus N315"				
	/strain="N315"				
	/sub_species="aureus N315"				
	/db_xref="taxon:158879"				
	complement(357..1205)				
	/gene="SA1571"				
	complement(357..1205)				
	/gene="SA1571"				
	/note="ORFID:SA1571"				
	/codon_start=1				
	/transl_table=1				
	/product="D-alanine aminotransferase"				
	/protein_id="BAB42839.1"				
	/db_xref="GI:13701546"				
	/translation="MEKIFLNGFVSPEAKVSYNDRCYVRCGCIYEIRYNNKLFPTVEIHERLRSANETGDLNLSVEELFELSKLVDNMOIETGALITYQATKGAARNHSFKYVPEVEAIVAYTKSYDRPYDLHENGNGVTVDIRMLRCDISLNLGNVLAKKEDNVAKYNAVEAIOHRGETVTEGSSNAAYVDGYIYTHPINNYILNCTIRTVIKKIDEDNIPFEEFTVDFLKNADDEVIVSTSAEVTVPVIKIDGPIINDGKGPITRQLQEGFEKXIESHSI"				
	complement(1209..2618)				
	/gene="SA1572"				
	complement(1209..2618)				
	/gene="SA1572"				
	/note="ORFID:SA1572"				
	kaa-His dipeptidase homolog"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="BAB42840.1"				
	/db_xref="GI:13701547"				
	/translation="MMKEKVOAQEDQIINDLGLAIESVVDAAKASADAPVGGPRRALADWVEIAHRDGFTHDVIHAGIEAGKNDVGLICHVDVYPAAGGMSNPFEPALVTEAIIARGLDLDKGPITIAIYVAIKILEDNANWKKRIIHAIITGDESDDKCRIDRYKTEMRPTLGRAPDAEFCIHEKGIITFDLVQNKLTEDQDPDELLITFKSGEIRNMV				

[illegible]

CDS

/gene="SA1577"
complement(7760..14320)
/gene="SA1577"
/note="ORFID:SA1577
hypothetical protein, similar to FntB protein"
/codon_start=1
/transl_table=11
/protein_id="BAB42845.1"
/db_xref="GI:13701552"
/translation="MNLKKKYSIRKYYGIFSTLIGVLLISNPGQAQTNDNV
QSDPNQATPVNSODTIVANRGLANSQNTPPNSATNTSTNOALVNRHNGSIANTAT
PISVQSTPSAQNNTHTDGTATETVSNANNNDVSNNTLVPKTKTENSGSGHLY
LKEIQEDVSRSSDKPELVIAEQASNPKRKRRAAPADPEAPAPAAAGANGAP
VAITAPPTPTDNNANAGQANREYSEFDNCRIRSTNSVSATVVDLPETTLIN
GKAVGFSHAMVRTSMFSGDANKYQOGVAVLIGRKNGTDHGDGVEIKSLVFN
SNELIEFEFTMTKTYOGVYVNLIRKADVDYIAKSVYGPIMLPFKPREVSHLK
LOFVPRKNDATDAGITQLRDQKYYVDVNSIGLHSGSHYVRRTRMEPRANKKEFT
VITSLKNGFGASFNTDFTVYQVLPGEVYVNSLTKDPSNSGVDMKDNVITVD
AANRVITIKSTGGSGNSPARLPMDKILDKYRLRVNVRLOQADYTPASDIDFKLRQOT
DFINSPAESHTVSTNPYTDIIIMKKDLQEVNRRIQADYTPASDIDFKLRQOT
ILDEKRNVPILNKRKSGQADISLNGQHTLIRSVDAENAVNKKVQOMEDLVNQNEL
TDEKKAQIOVIEBKHEIIGNIGDOTTGVRIRKDOGIGTISGDTAPVYVPMNKK
AIDKAKOREIINAPPADEDEIOLALNOLADERDADIDNTNATTNADVEIAKNG
INTIGAVVPOVTHKQARDAINQNTATKROQINSRPADEERNAALNELQVFNAL
EQINQATNDVDNAKGDGLAINPIAPVYVQAAADAVSHDAQHIAEINAMPAT
QEEKROAIDRVNAVTAANTNIIANNTNADVEQVNAIGQIAIPATVYKTDANA
IDKSAETQHTIFNNNDATLEEOQAAQQLDQAVATKQNIINAADTNOEVAQKDOG
ONTVIOPTOVGTDAKNAVNDKAREAITNINMPGATREKQEIINRVVITLKNRALT
DIGVSTTAMVNSIRDDAVNOIGAVOPHYVKQATCVLDLAFKAKOELNONTNAT
BEKQVALNOVDQELATINININQADTNAEDQAAQULCTKATINAIQPIYKAPALQOT
NOHNAKLAEINATPDATNDEKNAINTLQDQKQALLESIKQANTAEVQAAVTAEN
NIDAVQDVVAKQARDAIKITAEVAKRLEAVKQTPNATDEKQAAVNOINDKQDAINO
INQONTQVDTTQNAVNAIDNVEAVETKPAIADIEKAVEKCOQDINSIDSTDN
EKEVASQALKEKKEKALAAIDQATQNTSOVNAATNGVSAIKIKIIOPEKRVAPAREKIN
OKANEIAKINQKAEAREVQALDKINEFVNAQADITNNNTNOQVDTTQSOADS
IALVAPETVRAARDAVKKQTEKKEIQAQAHADDEEQAVALNOANNEKALQNTI
NQATNNDVREVETNGIATLKVOPHIVIPKQQAQKATAEQVESIKTPHATVDE
LDEANQILISLQAOQOEIENTQDAVTVNQTQIKAIQIIPKVRKKAALDSIEE
NNKQDLAINTLDTQDERTDAIDTLNKNIVNTKNDIAQNTNAEDVPTREGNNIN
KVLIPKVOVPAARQSVYKAEQANALIDQSDSTEEREAALHVLVQANQALIDQIN
HADTQAVNODSIDAQNVIKIRPATVYKPAQOQINAKTINKINLIKANKEATDEQ
NIAIAOYEKELIKAKQIASAVTNADVAYLHBEKKEIRIEVYINKKASARQULTL
FNDKQAIEANIQATVEERNSILAQNTQYDTAIGQIDORSNAQVDTKSLMQLIHD
DLVDHPKPKDAEKTINDDLARVATLQVNRKSNRKAALATIAITLQMDIELKT
ARTNADVAVALKRENVALSDIEAVITKEKSLIRIDIAQATKAFKAIATPOLKAV
KVLIDQVADGNRMIDEDATLNDIKQHTQPIVEIILAIKIPATATVSPKEIOPAPRV
CTPIKKEETHESRVEKELPNTGSEGDPLKEPALTGAALLARRRTKESK"

gene
complement(14646..14957)
/gene="SA1578"
complement(14646..14957)
/gene="SA1578"
/note="ORFID:SA1578"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42846.1"
/db_xref="GI:13701553"
/translation="MKSITVDELKKNLFESKPVQIVDVRTDETFANGYIPNATLIPMD

alignment_scores:
Quality: 715.00 Length: 136
Ratio: 5.257 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-728-670-10 x AP003135/rev ..

Align seg 1/1 to reverse of: AP003135 from: 1 to: 291150

1 SerSerSerPheaspLysgLySryLySgLyAspAspAlaSerThy 17
|||||
208296 TCAGTTCATTGCACAAAGAAATATATAAAGGCGATGACGCGATT 208247

17 rPhegluProThrglyProThryLeuMetValAsnValThrglyValAsps 34
|||||
208246 TTTTGAACCAACAGAGCCCGTATTGTGATGTAAGATGCTACGAGTTGATA 208197

34 erLySgLyAsnGluLeuLeuSerProHstYrValGluPheProIleLys 50
|||||
208196 GTAAAGGAAAGAAATGCTATCCCTCATATATGCGAGTTCCTATTTAA 208147

51 ProGlyThrThrLeuThrLySgLyLysIleGluThrYrValGluTyrPAL 67
|||||
208146 CCTGGACTACCTACTACAAAGAAAATTTGAATACATGCTCGAATGGC 208097

67 aLeuAspAlaThraLayrLySgLyGluPheArgValIleGluLeuAspPros 84
|||||
208096 ATTAGATCCGACACATATATAAGAGTTTAGAGTCTGGAATTAGATCCAA 208047

84 erAlaLySgIleGluValThrYrTyraSplysAsnLySgLyGluGlu 100
|||||
208046 GCGCAAGATCGAAGTCACCTATATATGATTAAGAAAGAAAAGAGAA 207997

101 ThrLySerPheProIleThrglyLySgLyPheValIleProAspLeuSe 117
|||||
207996 ACGAAGTCTTCCCTATACAGAAAAGCTTTTGTGTCCACAGATTATC 207947

117 rGluHisIleLySasnProGlyPheAsnLeuIleThrLySValIleG 134
|||||
207946 AGACATATTATAAAACCTGATTCACCTTAATTACAAAGTTGTTATAG 207897

134 LuLySlys 136
|||||
207896 AAAAGAA 207889

seq_name: gb_pat:A17530

seq_documentation_block:
LOCUS A17530 414 bp DNA linear PAT 12-APR-1994
DEFINITION Bacteriophage phi-C recombinant DNA for staphylokinase C-phi-C.
ACCESSION A17530
VERSION A17530.1 GI:512331
KEYWORDS staphylokinase.
SOURCE Bacteriophage phi-C.
ORGANISM Bacteriophage phi-C.
REFERENCE 1 (bases 1 to 414)
VIRUSES: dsDNA viruses, no RNA stage: Caudovirales; Siphoviridae.
AUTHORS
TITLE
JOURNAL
PATENT: WO 9313209-A 2 08-JUL-1993;
FEATURES
source
1. 414
/organism="Bacteriophage phi-C"
/specific_host="Staphylococcus aureus"
/db_xref="taxon:35242"
1. 414
/codon_start=1
/transl_table=11
/product="staphylokinase SAK-C-phi-C"
/protein_id="CAA01336.1"
/db_xref="GI:512332"
/translation="MSSEFDKGRKKYKGDASDFEPTGPIYLYVNTGVDGKGNELSPH
YVEPIKPGTTLTKERIEYVEMALDARAKERFVEELDPSAKIEVTYIDNKKKEET
KSPITKEKGFVVDLSEHIKNPGLNITKIVIEKR"
mat_peptide
4. 411
/product="staphylokinase SAK-C-phi-C"
BASE COUNT 161 a 59 c 83 g 111 t
ORIGIN

alignment_scores:
Quality: 711.00 Length: 136
Ratio: 5.228 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.265

alignment_block:

US-09-728-670-10 x A17530 .

Align seg 1/1 to: A17530 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysLysLysLysGlyAspAspLysSerTy 17
4 TCAGTTCATTCCGACAAAGGAAATATATAAAGGCGCATACCGCACTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
54 TTTTGAACCAACAAGGCCCGCTAATTTGGATGCTAAATGTGACCTGCAAGTTGATG 103
34 eLysGlyValAsnGluLeuLeuSerProHisLysTyrValGluPheProIleLys 50
104 GTAAAGCAAAATGAATTTGCTATCCCTCATTTATGTGCGAGTTTCTTATTAAA 155
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPAl 67
154 CTTGGGACTACACTTACAAAAGAAAAAATTGAATACTATCTCGAATGGGC 203
67 AlcuAspAlaThrAlaLeuTyrLysGluPheAlaGlyValGluLeuAspProS 84
204 ATTAGATGCCAGACACCATATAAAGGTTTAAAGTATGAAATTGATCCAA 255
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
254 GCCCAAGGATCGAAGTCACTTATATATATGAAGATAAGAAAAAGAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValIValProAspLeuSe 117
304 ACGAAGCTTTTCCCTATTAACAGAAAGAGGTTTGTGTTCACAGATTATC 355
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
354 AGAGCATATTTAAAAACCGTGAATCACTTAATTTACAAAGGTGTATTAG 403
134 LuLysLys 136
404 AAAAGAA 411

```

seq_name: gb_pat:AR037170

seq_documentation_block:			
LOCUS	AR0371170	414 bp	DNA
DEFINITION	Sequence 3 from patent US 5801037.		
FEATURES			
ORIGIN	1		

```
alignment_scores:
  Quality: 711.00
  Ratio: 5.228
  Gaps: 0
Percent Similarity: 100.000
Percent Identity: 99.265
```

```
alignment_block:
```

US-09-728-670-10 x AR037170 . .

Align seg 1/1 to: AR037170 from: 1 to: 414

1 SerSerSerPheaspIysGlyLysTyrLysLysGlyaspaspAlaSerTy 177

4 TCAAGTTCATTGCGACAAAGGAAAAATATATAAAAAGGGCCGATGACCGGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
54 TTTTACACCAACAGGCCCGGTATTTGATGCTAAATGTCACGTGAGATTGATG 103
34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
104 GTAAGGAAATGAATGCTATTCCTCCCTCATTAATGCGATTTCCATTTAAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
154 CCTGGAGCTACACTTACAAAAGAAAAAATGAATNCATATGCGAATGGGC 203
67 eLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
204 ATTATATGGCAGCATATTAAGATTAAAGATTAGGTGTAATTAATCAAA 253
84 eAlaLysIleGluValAlaThrTyrTyrAspLysAsnLysLysGluGlu 100
254 GGGCAAGAGTCACAACTCCTTATTTATGATTAAGAATAGAAAAAGAAAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValAlaProAspLeuS 117
304 ACGAAGTCTTCCATATACAGAAAAAAGGTTTGTGTGCCAGATTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
354 ACAGCATATTTAAAAACCTCGATTCACCTTAATTACAAAGGTTGTTATAG 403
134 LysLysLys 136
404 AAAAGCAA 411

seq_name: gb_pat:E02873

seq documentation block:			
LOCUS	E02873	804 bp	DNA
DEFINITION	DNA encoding staphylokinase(Sak).		
ACCESSION	E02873		
VERSION	E02873.1	GI:2171098	
KEYWORDS	JP 199108595-A/2.		
SOURCE	Staphylococcus aureus.		
ORGANISM	Staphylococcus aureus		

FT mat_peptide /note='SAK signal peptide'
 FT /product='394..801'
 FT /product='staphylokinase'
 FT /note='SAK'
 FT CDS 313..804
 FT /product='precursor of staphylokinase'.

FEATURES
 source Location/Qualifiers
 1..804
 /organism='Staphylococcus aureus'
 /db_xref='taxon:1280'

BASE COUNT 294 a 89 c 148 g 273 t
 ORIGIN

alignment_scores:
 Quality: 711.00 Length: 136
 Ratio: 5.228 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.265

alignment_block:
 US-09-728-670-10 x E02873 ..

Align seg 1/1 to: E02873 from: 1 to: 804

```

1 SerSerSerPheaspLysGlyTyrLysGlyaspaspalaSerty 17
|||||
394 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGATTA 443
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
444 TTTTGAACCAACAGCCCGCTATTGATGTAATGTGACTGAGTGTGATG 493
34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
494 GTAAAGGAATGATTCATTCCTATTCCTATGTCGAGTTTCTATTA 543
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrP 67
|||||
544 CTTGGGACTACACTTACAAAAGAAAATTTGAATGATGTCGATGGGC 593
67 aLeuaspalaThrAlaTyrLysGluPheArgValGluLeuaspProS 84
|||||
594 ATTAGATCGACAGCATATTAAGAGTTTAGAGTGTGAATTAAGATCAA 643
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
644 GCGCAAAAGATCGAAGTCACTTATATGATTAAGATTAAGAAAAGAA 693
101 ThrLysSerPheProIleThrGluLysGlyPheValProaspLeuSe 117
|||||
694 ACGAAGCTTTCCCTTACACAAAAGGTTTGTGTCAGATTATATC 743
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134
|||||
744 ACGACATATTAATAAACCTGATCACTTAATTAACAAGGTTGTTATG 793
134 LuLysLys 136
|||||
794 AAAAGAAA 801

```

seq_name: gb_ba:SPSAK1

seq_documentation_block:

LOCUS SPSAK1 1377 bp DNA linear BCT 30-MAR-1995

DEFINITION Staphylococcus aureus S-phi-C gene for staphylokinase.

VERSION X00127.1 GI:47425

KEYWORDS kinase; plasmidogen activator; signal peptide; staphylokinase.

SOURCE Staphylococcus aureus
 ORGANISM Staphylococcus aureus
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/staphylococcus group; Staphylococcus.
 REFERENCE 1 (bases 1 to 1377)

AUTHORS Sako, T. and Tsuchida, N.
 TITLE Nucleotide sequence of the staphylokinase gene from Staphylococcus aureus
 JOURNAL Nucleic Acids Res. 11 (22), 7679-7693 (1983)
 MEDLINE 84069795

FEATURES
 source Location/Qualifiers
 1..1377
 /organism='Staphylococcus aureus'
 /strain='phage S-phi-C'
 /db_xref='taxon:1280'

promoter 193..198
 /note='pos. promoter region'
 promoter 217..222
 /note='pos. promoter region'
 RBS 301..305
 /note='ribosome binding site'
 CDS 313..804
 /codon_start=1
 /transl_table=11
 /product='staphylokinase'
 /protein_id='CAA24957.1'
 /db_xref='GI:758303'

/db_xref='SWISS-PROT:P00802'
 /translation='MKRSLLFLVLLLFSSSITNEVSASSSFQKGYKKGDAST
 FEPPTPLAVNTGVDGKGNELSPHYVEFPKPGTTLTKERIEYVEMADATAYKE
 FRVVELDPSAKIEVTYDKNKKKEETKSFPIEKGFFVVDLSHIKNGFNLIIRKVI
 EKK'

sig_peptide 313..393
 /note='putative'
 mat_peptide 394..801
 /product='staphylokinase'
 BASE COUNT 452 a 184 c 255 g 486 t
 ORIGIN

alignment_scores:
 Quality: 711.00 Length: 136
 Ratio: 5.228 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.265

alignment_block:
 US-09-728-670-10 x SPSAK1 ..

Align seg 1/1 to: SPSAK1 from: 1 to: 1377

```

1 SerSerSerPheaspLysGlyTyrLysGlyaspaspalaSerty 17
|||||
394 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGATTA 443
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
444 TTTTGAACCAACAGCCCGCTATTGATGTAATGTGACTGAGTTTATG 493
34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
494 GTAAAGGAATGATTCATTCCTATTCCTATGTCGAGTTTCTATTA 543
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrP 67
|||||
544 CTTGGGACTACACTTACAAAAGAAAATTTGAATGATGTCGATGGGC 593
67 aLeuaspalaThrAlaTyrLysGluPheArgValGluLeuaspProS 84
|||||
594 ATTAGATCGACAGCATATTAAGAGTTTAGAGTGTGAATTAATCA 643
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
644 GCGCAAAAGATCGAAGTCACTTATATGATTAAGATTAAGAAAAGAA 693
101 ThrLysSerPheProIleThrGluLysGlyPheValProaspLeuSe 117
|||||
694 ACGAAGCTTTCCCTATTAACAAAAGGTTTGTGTCAGATTATATC 743

```

```

117 rgluhisiletyasapnProGlyPheasnLeuilethrLysvalValileg 134
|||||
744 AGAGCATATATAAAACCCGTGATTCACTTAATTAACAAAGTTGTATAG 793
134 JulysLys 136
|||||
794 AAAAGAAA 801

seq_name: gb_pat:AX073708

seq_documentation_block:
LOCUS AX073708 1377 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 1 from Patent WO0104287.
ACCESSION AX073708
VERSION AX073708.1 GI:12710128
KEYWORDS
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
1 (bases 1 to 1377)
AUTHORS Haliker,T., Pedersen,A.H. and Okkels,J.S.
TITLE A method for preparing modified polypeptides
JOURNAL Patent: WO 0104287-A 1 18-JAN-2001;
Maxygen Aps (DK)
FEATURES
source 1..1377
Location/Qualifiers
Promotor /organism="Staphylococcus aureus"
193..232 /db_xref="taxon:1280"
KBS 301..305
S19_peptide 313..393
BASE COUNT 452 a 184 c 255 g 486 t
ORIGIN

alignment_scores:
Quality: 711.00 Length: 136
Ratio: 5.228 Gaps: 0
Percent similarity: 100.000 Percent identity: 99.265

alignment_block:
US-09-728-670-10 x AX073708 ..

Align seq 1/1 to: AX073708 from: 1 to: 1377
1 SerSerSerPheAspLysGlyLysLysGlyAspAspLaseTyr 17
|||||
394 TCAAGTTCATTCGACAAAGCAAAATATAAAAGCGCATACCGAGCTTA 443
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
444 TTTTGAACCAACAGGCCGATATTGATGATGAATGTGATCGAGTGGATG 493
34 erysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
494 GTAAAGAAATGAATGTCTATCCCTCATATATGCGATTCCTATTAA 543
51 ProGlyThrThrLeuThrLysGlyLysIleGlyTyrTyrValGluTrrP 67
|||||
544 CTGTGCGACTACACTTACAAAAAGAAAATTGATACATATCCAAATGGC 593
67 AleuAspAlaThrAlaTyrLysGluPheArgValAlaGluLeuAspProS 84
|||||
594 ATTGATGCGCACACATATAAGAGTTTAAAGTAGTAGTGAATTGATGATCA 643
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
644 GCGGAAAGATCGAAGTCACTTATTATGATAGATAAGAAAAAGAGAA 693
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuS 117
|||||
694 ACGAAGCTTTCCCTATACAGAAAAAGTTTGTCTTCTCCACATTTATTC 743

```

```

117 rgluhisiletyasapnProGlyPheasnLeuilethrLysvalValileg 134
|||||
744 AGAGCATATATAAAACCCGTGATTCACTTAATTAACAAAGTTGTATAG 793
134 JulysLys 136
|||||
794 AAAAGAAA 801

seq_name: gb_pa:AP003364

seq_documentation_block:
LOCUS AP003364 348650 bp DNA linear BCT 07-FEB-2002
DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 7/9.
ACCESSION AP003364 BA000017
VERSION AP003364.2 GI:14247707
KEYWORDS
SOURCE Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
strain:Mu50) DNA.
Staphylococcus aureus subsp. aureus Mu50
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
1 (sites)
AUTHORS Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,
Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kato,C.,
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuuchi,J.,
Kanehisa,M., Yamashita,A., Oshima,S., Furuya,K., Yoshino,C.,
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiratsuka,T.
TITLE Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL Lancet 357 (9264), 1225-1240 (2001)
MEDLINE 2311952
REFERENCE
2 (bases 1 to 348650)
AUTHORS Ohta,T.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology, 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail:tohata@tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)
On May 29, 2001 this sequence version replaced gi:13873637.

COMMENT
FEATURES
source 1..348650
Location/Qualifiers
/organism="Staphylococcus aureus subsp. aureus Mu50"
/strain="Mu50"
/sub_species="aureus Mu50"
/db_xref="taxon:158878"
complement(59..235)
/gene="SAV1935"
complement(193..235)
/gene="SAV1935"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="BAB58097.1"
/db_xref="GI:14247708"
/translation="MKHLTKIFVLAIIIFIGYVLDATNHESGIGIKLLAIMEPIC
AFISRSNDRKRNK"
775..2061
/gene="SAV1936"
775..2061
/gene="SAV1936"
/note="similar to aspartate transaminase protein"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="BAB58098.1"
/db_xref="GI:14247709"
/translation="MNPLOSILNEOLQGSNATFTMLSDLGOMPFYKGLISQSAEAK
STIYNATIGATNKDKMKFASISLDAPFNDLTPETIPYAPPOSIEELRDLMOQKMLRD

```



```

gene
CDS
    complement(2340..4075)
    /gene="SAV1937"
    /note="SAV1937"
    /codon_start=1
    /transl_table=11
    /product="truncated map-w protein"
    /protein_id="BAB58099.1"
    /db_xref="GI:14247710"
    /translation="MKKKKSNVKKPEPTINKENITVSEFVFNASKRTITDLSIK
    LKSNMADQGITKHDIQLAEKAVKVFYFKNGSKYDVKTEKDERVFKATDIKKVDI
    ELKF"
    complement(2645..4075)
    /gene="SAV1938"
    /note="SAV1938"
    /codon_start=1
    /transl_table=11
    /product="truncated map-w protein"
    /protein_id="BAB58100.1"
    /db_xref="GI:14247711"
    /translation="MKFKSLITTLALGVASNGANTNEASAAKPLDKSSTLIH
    GHSNIOIPYITVNGTSONLSLTFENKNONISYKDEENKVSYLENRSIDDLRL
    SKQAEYVHERKNGKRYIDLSGITYTDLINTSDIKAISVNDTKKOPKRAKANOV
    PFTTVNGTSONLSLTFENKNONISYKDEENKVSYLENRSIDDLRLSKQAKT
    VNFNGTKKVIDLSGITYTANLNSSDIKSININVDPKHLENKAKRNOVPSINIG
    GTSTNLSNLSFNSKRPMTNKNLTSQIKSVLKHDRGISEODLKYAKKAYTVFKNG
    KRIQLNSKNYATLVAKDYKRIEITVKTGAKADRYVPTIYAVNGSTPLSTK
    ISNQLISYKINDKVSYLENRSIDDLKFAKQAKYVYFKNGKQVYNLKSDF
    TPNLFSAKDIKKIDIVKQYTKSKK"
    4418..4660
    /gene="truncated-hlb"
    4418..4660
    /gene="truncated-hlb"
    /note="SAV1939"
    /codon_start=1
    /transl_table=11
    /product="truncated 6449-6450 beta-hemolysin"
    /protein_id="BAB58101.1"
    /db_xref="GI:14247712"
    /translation="MLHLFCCLYKRSNDGKKTKKSLKAVTALANLLVGLTD
    NSAKAESKKDDTDLKLVSNNYMLSTVLYFRWRLLT"
    5032..5211
    /gene="SAV1940"
    5032..5211
    /gene="SAV1940"
    /note="Bacteriophage phi1315"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAB58102.1"
    /db_xref="GI:14247713"
    /translation="MKFKTKRECKSNINFRSQEINNSESEKGLWGISMLILFLLI
    LFGITACSSSHIFIN"
    5370..5543
    /gene="SAV1941"
    5370..5543
    /gene="SAV1941"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAB58103.1"
    /db_xref="GI:14247714"
    /translation="MIGIFVAIVGKRAVYVLLNLLIFTNIFVFLHFCVLLLKYGIF
    KLFYEGCVLLFT"
    complement(5834..6184)
    /gene="SAV1942"
    gene

```

```

CDS
    complement(5834..6184)
    /gene="SAV1942"
    /note="Bacteriophage phi1315"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAB58104.1"
    /db_xref="GI:14247715"
    /translation="MKIRSLIAGTAVIASPLVTNLDKNEAQSSTLPTSNEXONE
    KIANELKSLDELAVNELATGSLNYYRTIKISQKAMVYALKSDFKMSKAYQLD
    KTYNEIDELAKSKY"
    complement(6695..7162)
    /gene="truncated-lytA"
    complement(6695..7162)
    /gene="truncated-lytA"
    /note="Bacteriophage phi1315"
    SAV1943"
    /codon_start=1
    /transl_table=11
    /product="truncated amidase"
    /protein_id="BAB58105.1"
    /db_xref="GI:14247716"
    /translation="MLAIDILFALSHISQOHRNITVATPHLKCGYFYKCVSGIS
    NKTPAKNENSPVPAGITLDKNVVPYKKEITVYVANKGNVNDGISTNSRTIGVL
    PNNATIKYDGAICINGYRWITVYIANNQRCYIATGEVDKAGNRISSEFNSAL"
    complement(7682..8173)
    /gene="sak"
    complement(7682..8173)
    /gene="sak"
    /note="Bacteriophage phi1315"
    SAV1944"
    /codon_start=1
    /transl_table=11
    /product="staphylokinase precursor"
    /protein_id="BAB58106.1"
    /db_xref="GI:14247717"
    /translation="MLKRSLLFLVYVLLLSFSSITNEVSASSSFDKRYKKGDASY
    FEPTGPLYAVNVYGVGDKGNELSPHYVEFPYKPGTTLKEKIEYVEMALDAPAYKE
    FRVVELDPSAKIEVYYDDKNNKKEKSPPIYKGFVVDLSSEHIKNGFNLITKVI
    EKK"
    complement(8388..9119)
    /gene="SAV1945"
    complement(8388..9119)
    /gene="SAV1945"
    /note="Bacteriophage phi1315"
    /codon_start=1
    /transl_table=11
    /product="lytic enzyme"
    /protein_id="BAB58107.1"
    /db_xref="GI:14247718"
    /translation="MKTYSEARARLRWYGRYIDPFGWYGYOCADLAVDYIYMLEIR
    MGNAKDAINDENKNAIVYENTPSFVPIGVAVFTGIVGIGHGIVGNGTNO
    FLILBNYDGNANTPAKLRWDNYGCTGTHIRKRYSEGLMNTIRKVPRAOKANGS
    ASKITVGSKAPYILKWSKGAIFNAKIDGLGATSAIRYGNRTNRFVGOAVYAFGLT
    IYVELIIDWCRIYNNHNMWICYE"
    complement(9131..9385)
    /gene="SAV1946"
    complement(9131..9385)
    /gene="SAV1946"
    /note="Bacteriophage phi1315"
    /codon_start=1
    /transl_table=11
    /product="holin homolog"
    /protein_id="BAB58108.1"
    /db_xref="GI:14247719"
    /translation="MINMKIRMKQKSFVWIIISALFLFQNTAKAIGDIOVYTBQLT
    alignment_scores:
    Quality: 711.00 Length: 136
    Ratio: 5.228 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 99.265
    alignment_block:

```


CDs
NPDLIVIDEALSVGDQTFPAOKCDKIEYEKEONKTFVSHNLGVROFCTKIAMI
GKLDYGELODVLPRKKEAFLENDKRRSKAEQKRFKRLDEARFVIK*
11017. .12066
/codon_start=1
/transl_table=11
/product="tag homolog"
/protein_id="BA95014.1"
/db_xref="GI:7670327"
/translation="MSAIGTVEKEHKNELLQRLAOFQVKIINSHSYLGAVAMELINP
VMOIYVMVFGIGISNAPIHGVFVYMLVIGISMFEINOGILEGTRAKIOTKNOY
SKNFPISITPIYTSRFYGHGILLIYVACMFYISPIHIOILLIYVPCFILT
ASVTLSTLIGVLRDPMOMQAILILYFSTILMPKNGISGLIHEMKYNPYE
IASISRAILIHETFMDSALVILVAPINOTNGSEADQIPSGYTIATGTDPDVH
NTTIRPDYVFIPIPVGQIRAVVYVYNNKVVASMSNAVTIIPDIPPTINNPGINA
KYYRDEVNFTMGVC"

BASE COUNT 5264 a 2419 c 2914 g 5003 t
ORIGIN

alignment_scores:
Quality: 710.00 Length: 136
Ratio: 5.221 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.529

alignment_block:
US-09-728-670-10 x AB033232 ..
Align seg 1/1 to: AB033232 from: 1 to: 15600

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
6529 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGATTA 6578
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
6579 TTTTGAACCAACAGGCCCGCTATTGATGTGAATGTGACTGAGTGATG 6628
34 eTyrGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|||||
6629 GTAAAGGAATGAAATGCTATCCCTCATTTATGAGACTTCCATTATAA 6678
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrAl 67
|||||
6679 CTTGGGACTACCTTACAAAGAAATAATGAACTATGTCGAATGGGC 6728
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
6729 ATTAGATGGCAGACATATAAGAGTTTAGACTTGAATTAGATCCAA 6778
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
6779 GCGCAAGATCGAAGTCACTTATGATTAAGATAAGAAAAAGAAAGAA 6828
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
6829 ACGAAGCTCTTCCCTATTAACAGAAAAAGTTTGTTCGCCAGATTATC 6878
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
6879 AGAGCATATTAAAAACCTCGATTCACCTTATTACAAAAGTTATTATAG 6928
134 LuLysLys 136
|||||
6929 AAAAGAAA 6936

seq_name: gb_Da:AF332619
seq_documentation_block:
LOCUS AF332619 727 bp DNA linear BCT 21-FEB-2001
DEFINITION Staphylococcus aureus staphylokinase SakXh gene, complete cds.
ACCESSION AF332619
VERSION AF332619.1 GI:13022096
KEYWORDS

SOURCE
ORGANISM
Staphylococcus aureus.
Staphylococcus aureus
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
1 (bases 1 to 727)
AUTHORS
Wei, W., Xiang, H. and Tan, H.
TITLE
Direct Submission
JOURNAL
Submitted (28-DEC-2000) Genetics, Institute of Microbiology,
Chinese Academy of Sciences, 13, Belyitiao, Zhongguancun, Beijing
100080, P. R. China

FEATURES
source
1. .727
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
RBS
224. .228
236. .727
CDS
236. .727
/function="catalyzes plasminogen to plasmin"
/note="protein kinase; similar to staphylokinase variants
sak42D, sakSTAR and sakC"
/codon_start=1
/transl_table=11
/product="staphylokinase SakXh"
/protein_id="AKK11636.1"
/db_xref="GI:13022097"
/translation="MLKRSLLFLVLLLFSSITNEVSASSSPDKYKKGDASY
FEPTGPYLMVNTGVGDKGNELISPRVYEPFIKPTLTLEKEIEYVEMALDAPYKE
FRVVELDPSAKIEVYYDDKNKKKEETKSFPITEKGFVVPDLSHIKNGFNLIRVVI
EKK"

BASE COUNT 269 a 80 c 132 g 246 t
ORIGIN

alignment_scores:
Quality: 703.00 Length: 136
Ratio: 5.169 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.529

alignment_block:
US-09-728-670-10 x AF332619 ..
Align seg 1/1 to: AF332619 from: 1 to: 727

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
317 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGATTA 366
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
367 TTTTGAACCAACAGGCCCGCTATTGATGTGAATGTGACTGAGTTGATG 416
34 eTyrGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|||||
417 GTAAAGGAATGAAATGCTATCCCTCGTATGTCGAGTTTCCATTATAA 466
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrAl 67
|||||
467 CTTGGGACTACCTTACAAAGAAATAATGAACTATGTCGAATGGGC 516
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
517 ATTAGATGGCAGACATATAAGAGTTTAGAGTTGAAATTTAGATCCAA 566
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
567 GCGCAAGATCGAAGTCACTTATTATGATAAGATAAGAAAAAGAAAGAA 616
617 ACGAAGCTCTTCCCTATTAACAGAAAAAGTTTGTTCGCCAGATTATC 666
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
667 AGAGCATATTAAAAACCTCGATTCACCTTAAATTAACAAAAGTTGTTATAG 716

134 luluyslys 136
|||||
717 AAAAGAAA 724

seq_name: gb_ba:SAU77328

seq_documentation_block:

LOCUS SAU77328 411 bp DNA linear BCT 14-JUN-1999
DEFINITION Staphylococcus aureus staphylokinase gene, partial cds.
ACCESSION U77328
VERSION U77328.1 GI:2605637

KEYWORDS

SOURCE Staphylococcus aureus.

ORGANISM Staphylococcus aureus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.

REFERENCE 1 (bases 1 to 411)

AUTHORS Kim, S.H., Chun, H.S., Han, M.H., Park, N.Y. and Suk, K.
TITLE A novel variant of staphylokinase gene from Staphylococcus aureus
ATCC 29213

JOURNAL Thromb. Res. 87 (4), 387-395 (1997)

MEDLINE 97417758

PUBMED 9271816

REFERENCE 2 (bases 1 to 411)

AUTHORS Chun, H.S., Suk, K. and Kim, S.H.
TITLE Direct Submission

JOURNAL Submitted (04-NOV-1996) Protein Function R.U., KRIBB, KIST, PO Box 115, Yusong, Taejeon 305-600, Korea

FEATURES

source 1..411
/organism="Staphylococcus aureus"
/strain="ATCC 29213"
/db_xref="taxon:1280"
CDS <1..411
/codon_start=1
/product="staphylokinase"
/protein_id="AAB84174.1"
/db_xref="GI:2605638"
/translation="SSSPDKGKYYKKGDADSYPEPTGPIYLVNVTGVGDKRNELLSPHYVEFPKPGTTLTKKIEIYVEMALDATALYKERRVVELDPSAKIEVTYYDKKKKEETKSEPIREKGFVVDLSEHKNPENLTIKVIEKK"

BASE COUNT 162 a 59 c 81 g 109 t
ORIGIN

alignment_scores:
Quality: 699.00 Length: 136
Ratio: 5.178 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:
US-09-728-670-10 x SAU77328 ..

Align seg 1/1 to: SAU77328 from: 1 to: 411

```

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
1 TCAGATTCATTCGACAAAGGAAATATATMAAAGCGGATGACCGAGTTA 50
|||||
17 rPhgGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
51 TTTTGAACCAACAGCGCCGTAATTGATGTAATGTGACGAGGATGAGG 100
|||||
34 eRLysGlyAenGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
101 GTAAAGAAATGAATGCTATCCCATATATGCGAGTTCTCTATATAA 150
|||||
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrAl 67
|||||
151 CTTGGGACTACATTACAAAAAATTAATGAAATGACTATGTCGAAATGGGC 200
|||||

```

seq_name: gb_pat:A17529

seq_documentation_block:

LOCUS A17529 414 bp DNA linear PAT 12-APR-1994
DEFINITION Bacteriophage 42D recombinant DNA for plasmidogen activator SAK.
ACCESSION A17529
VERSION A17529.1 GI:490010

KEYWORDS staphylokinase.
Bacteriophage 42D.

SOURCE Bacteriophage 42D

ORGANISM Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
Lambda phage group.
1 (bases 1 to 414)

REFERENCE 1

AUTHORS EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES

JOURNAL Patent: WO 9313209-A 1 08-JUL-1993;
FEATURES Location/Qualifiers

source 1..414

CDS 1..414
/organism="Bacteriophage 42D"
/specific_host="Staphylococcus aureus"
/db_xref="taxon:10715"
1..414
/codon_start=1
/transl_table=1
/product="staphylokinase SAK42D"
/protein_id="CAA01335.1"
/db_xref="GI:490011"
/translation="MSSSPDKGKYYKKGDADSYPEPTGPIYLVNVTGVGDKRNELLSPHYVEFPKPGTTLTKKIEIYVEMALDATALYKERRVVELDPSAKIEVTYYDKKKKEETKSEPIREKGFVVDLSEHKNPENLTIKVIEKK"

mat_peptide 4..411
/product="staphylokinase SAK42D"
BASE COUNT 161 a 59 c 82 g 111 t 1 others
ORIGIN

alignment_scores:
Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:
US-09-728-670-10 x A17529 ..

Align seg 1/1 to: A17529 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
4 TCAGATTCATTCGACAAAGGAAATATMAAAGCGGATGACCGAGTTA 53
|||||
17 rPhgGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTGACCAACAGCGCCGTAATTGATGTAATGTGACGAGTTGATG 103
|||||

```

```

34 ertLysGlyAsnGluLeuLeuSerProHistiTyValGluPheProIleLys 50
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 GTAAAGAAATGAATTCCTATCCCTGTTATGTCGACTTCCATTTAA 153

51 ProGlyThrThrLeuThrLysGluLysIleGluTyTyValGluTPrAl 67
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 CTTGGAGTACACTTACAAAGAAAAAATGAACTGATGTCGAAATGGGC 203

67 aleuAspAlaThrAlaTyTyLysGluPheArgValGluLeuAspProS 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 ATTAAATGCGACAGCATATTAAGAGTTAGAGTACTGATTAATGATCCAA 253

84 ertAlaLysIleGluValThrTyTyAspLysAsnLysLysGluGlu 100
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 GCGCAAGATCGAAGTCACTTATATGATTAAGAAATGAAAAAGAGAA 303

101 ThrLysSerPheProIleThrGluLysGlyPheValAlProAspLeu 117
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 AGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCCAGATTATTC 353

117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 AGAGCATATTAACAACTGATTAACCTTAATTAACNAGTTGTTATAG 403

134 LuLysLys 136
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 AAAAGAAA 411

seq_name: gb_pat:AR037169

seq_documentation_block:
LOCUS AR037169 414 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5801037.
ACCESSION AR037169
VERSION AR037169.1 GI:5955025
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 414)
AUTHORS Behnke,D., Schloft,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.
TITLE Expression of signal-peptide-free staphylokinases
JOURNAL Patent: US 5801037-A 1 01-SEP-1998;
FEATURES
source 1..414
location/Qualifiers
BASE COUNT 161 a 59 c 82 g 111 t 1 others
ORIGIN

alignment_scores:
Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:
US-09-728-670-10 x AR037169 ..

Align seg 1/1 to: AR037169 from: 1 to: 414
1 SerSerSerPheAspLysGlyLysTyTyLysLysGlyAspAspAlaSerTy 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAGCGCATGACGAGTTA 53
17 rPheGluProThrGlyProTyTyLeuMetValAsnValThrGlyValAspS 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 TTTTGAACCAACAGCGCCGCTATTTGATGTAATGATGACTGAGTTGATG 103

34 ertLysGlyAsnGluLeuLeuSerProHistiTyValGluPheProIleLys 50
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 GTAAAGAAATGAATTCCTATCCCTGTTATGTCGACTTCCATTTAA 153

```

```

51 ProGlyThrThrLeuThrLysGluLysIleGluTyTyValGluTPrAl 67
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 CTTGGAGTACACTTACAAAGAAAAAATGAACTGATGTCGAAATGGGC 203

67 aleuAspAlaThrAlaTyTyLysGluPheArgValGluLeuAspProS 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 ATTAAATGCGACAGCATATTAAGAGTTAGAGTACTGATTAATGATCCAA 253

84 ertAlaLysIleGluValThrTyTyAspLysAsnLysLysGluGlu 100
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 GCGCAAGATCGAAGTCACTTATATGATTAAGAAATGAAAAAGAGAA 303

101 ThrLysSerPheProIleThrGluLysGlyPheValAlProAspLeu 117
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 AGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCCAGATTATTC 353

117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 AGAGCATATTAACAACTGATTAACCTTAATTAACNAGTTGTTATAG 403

134 LuLysLys 136
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 AAAAGAAA 411

seq_name: gb_pat:AX247545

seq_documentation_block:
LOCUS AX247545 534 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 21 from Patent WO0166776.
ACCESSION AX247545
VERSION AX247545.1 GI:15862238
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 534)
AUTHORS Holtschen,C., Gumpert,J., Kujau,J.M., Fritsche,C., Elske,G.,
Fahner,T.B., Sieben,S. and Mueller,H.P.
TITLE Novel 1-form bacterial strains, method for producing same and the
use thereof for producing gene products
JOURNAL Patent: WO 0166776-A 21 13-SEP-2001;
FEATURES
source 1..534
location/Qualifiers
CDS
1..534
/organism="Proteus mirabilis"
/db_xref="taxon:584"
/note="CemaH1-Sak"
/codon_start=11
/transl_table=11
/db_xref="GI:15862239"
/translation="MDNKRQDIDIFSIITIIWSCALMAFNRYQIAGESTIREQSS
FKGKYKKGDSDSYEPGPYLMVNTVGDGRNRLSPRVVEPIKGTTLTKRKIE
YVEWALDATAYKEFRVVELDPSAKIEVTYDKNKKEETKSFPTTEGFFVPLSEH
IKNPGNLTITKVIEKK"
BASE COUNT 201 a 77 c 111 g 145 t
ORIGIN

alignment_scores:
Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:
US-09-728-670-10 x AX247545 ..

Align seg 1/1 to: AX247545 from: 1 to: 534
1 SerSerSerPheAspLysGlyLysTyTyLysLysGlyAspAspAlaSerTy 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

124 TCAGTTCATTGCGAAGAAATATATAAAAGCGATGACGGAGTTA 173
17 rphglupProthrGlyProtyrLeuMetValAsnValThrGlyValAsp 34
|||||
174 TTTTGAACCAAGACGCGCGTATTTGATGTAATGTCACGAGTTGATG 223
34 erlysgLyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|
224 GTAAAGAAATGAAATGCTATCCCTGTTATGCGAGTTTCCATTTAA 273
51 ProGlyThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPAL 67
|||||
274 CCTGGGACTCACTTACANNAAGAAAATTGAATCTATGTCGATGGC 323
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
324 ATTAGATGCGACAGCATATTAAGAGTTTAGAGTATGAATTTAGATCAA 373
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
374 GCGCAAGATCGAAGTCATTTATGATTAAGATAAGAAAAAGAAAGAA 423
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
424 ACAGAGCTTCCCTATACAGAAAAGGTTTGTGTCCAGATTATATC 473
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
474 AGAGCATATTTAAAAACCTGATTCACCTTAATTAACAAGGTTGTATAG 523
134 LuLysLys 136
|||||
524 AAAAGAAA 531
seq_name: gb_pat:AX247537

seq_documentation_block:
LOCUS AX247537 540 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 13 from Patent WO0166776.
ACCESSION AX247537
VERSION AX247537.1 GI:15862230
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 540)
AUTHORS Holtschen,C., Gumpert,J., Kujau,J.M., Fritsche,C., Elske,G.,
Fahner,B., Sieben,S. and Mueller,H.P.
TITLE Novel 1-form bacterial strains, method for producing same and the
use thereof for producing gene products
PATENT: WO 0166776-A 13 13-SEP-2001;
JOURNAL FUER MOLEKULARE BIOLOGIE E.V. (DE)
INSTITUT location/Qualifiers
FEATURES
source 1..540
/organism="Escherichia coli"
/db_xref="taxon:562"
1..540
/contig="LacyH1-Sak"
/codon_start=1
/transl_table=11
/protein_id="CAC86585.1"
/db_xref="GI:15862231"
/translation="MYLKNTFMFMFGLFFPPFTLMGAYPPFPPLIHLHDINHISLOS
SFPDGKKYKKGDDASVFEPTGYLMVNVGVGKNNELLSPRVEPIKPGTTLREK
IEVYEMALDAATAYKEFRVLEIDPSAKIEVITYYDKNKKKEETKSPITPKGVVDLS
EIKKMGRLITKVIYEK"
BASE COUNT 192 a 87 c 99 g 162 t
ORIGIN
alignment_scores:
Quality: 695.00 Length: 136

```

```

Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794
alignment_block:
US-09-728-670-10 x AX247537 ..
Align seq 1/1 to: AX247537 from: 1 to: 540

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTyr 17
|||||
130 TCAGTTCATTGCGAAGAAATATATAAAAGCGATGACGGAGTTA 179
17 rphglupProthrGlyProtyrLeuMetValAsnValThrGlyValAsp 34
|||||
180 TTTTGAACCAAGACGCGCGTATTTGATGTAATGTCACGAGTTGATG 229
34 erlysgLyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|
230 GTAAAGAAATGAAATGCTATCCCTGTTATGCGAGTTTCCATTTAA 279
51 ProGlyThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPAL 67
|||||
280 CCTGGGACTCACTTACANNAAGAAAATTGAATCTATGTCGATGGC 329
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
330 ATTAGATGCGACAGCATATTAAGAGTTTAGAGTATGAATTTAGATCAA 379
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
380 GCGCAAGATCGAAGTCATTTATGATTAAGATAAGAAAAAGAAAGAA 429
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
430 ACAGAGCTTCCCTATACAGAAAAGGTTTGTGTCCAGATTATATC 479
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
480 AGAGCATATTTAAAAACCTGATTCACCTTAATTAACAAGGTTGTATAG 529
134 LuLysLys 136
|||||
530 AAAAGAAA 537
seq_name: gb_pat:AX247541

seq_documentation_block:
LOCUS AX247541 639 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 17 from Patent WO0166776.
ACCESSION AX247541
VERSION AX247541.1 GI:15862234
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 639)
AUTHORS Holtschen,C., Gumpert,J., Kujau,J.M., Fritsche,C., Elske,G.,
Fahner,B., Sieben,S. and Mueller,H.P.
TITLE Novel 1-form bacterial strains, method for producing same and the
use thereof for producing gene products
PATENT: WO 0166776-A 17 13-SEP-2001;
JOURNAL FUER MOLEKULARE BIOLOGIE E.V. (DE)
INSTITUT location/Qualifiers
FEATURES
source 1..639
/organism="Escherichia coli"
/db_xref="taxon:562"
1..639
/contig="SecAH1-Sak"
/codon_start=1
/transl_table=11
/protein_id="CAC86587.1"
/db_xref="GI:15862235"

```

/translation="MAKQPLDFQSAKGIGELKRRLLFVIGALLVFRIGSFIPRIGI
DAAVLAKLEQROGTIIEMFNMFSGALISRLQSSSPKGGKDDASVEEPYILM
VNTVGVGKRNELISPRYVEPIKGTLLKREKLEYEVALADATAYKEFRVVELPS
AKIEVYVDKNNKKEETKSPITEKGFVVDLSEHIKNPQNITIVIEKK"

BASE COUNT 209 a 110 c 141 g 179 t

ORIGIN

alignment_scores: Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:
US-09-728-670-10 x AX247541 ..

Align seg 1/1 to: AX247541 from: 1 to: 639

```

1 SerSerSerPheaspLysGlyLysTyLysGlyaspaspalaSerty 17
|||||
229 TCAGATTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGATTA 278
17 rPhegluProthrGlyProtyrLeuMetValasnValThrGlyValasps 34
|||||
279 TTTTGAACCAACAGCCCGCTATTGTGTAATGTGACTGAGGAGTTGANG 328
34 erLysGlyasnGluLeuSerProHisTyrValGluPheProileLys 50
|||||
329 GTTAAAGAAATGATTCATTCCTCCCTGATGCGAGTTCTTATTA 378
51 ProGlyThrThrLeuThrLysGlyLysIleGlyTyTyValGluTTPAI 67
|||||
379 CCGGCGACTACACTTACAAAAGAAAATTGAATCTATGTCGAATGGCG 428
67 aleuaspalaThrAlaTyryLysGluPheargValValGluLeuaspPro 84
|||||
429 ATTAGATCGACAGCATATTAAGAGTTTAGAGTAGTGAATTAGATCCAA 478
84 erAlaLysIleGluValThrTyTyTyAspLysAsnLysLysGlyGlu 100
|||||
479 GCGCAAAAGATCGAAGTCACTTATATGATTAAGATAAGAAAAGAAAGA 528
101 ThrLysSerPheProIleThrGlyLysGlyPheValProaspLeuSe 117
|||||
529 ACGAAGTCTTCCCTATTAACAGAAAAGGTTGTGTGCCAGATTATTC 578
117 rGluHisIleLysasnProGlyPheasnLeuIleThrLysValValIleG 134
|||||
579 AGAGCATATTAATAAACCTGATTCACCTTAATTAACAAAAGTTGTATAG 628
134 LuLysLys 136
|||||
629 AAAAGAAA 636

```

seq_name: gb_pat:AX247539

seq_documentation_block:

LOCUS AX247539 720 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 15 from Patent WO0166776.
ACCESSION AX247539
VERSION AX247539.1 GI:15862232
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Bacteria: Escherichia coli
Escherichia.
Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;

REFERENCE 1 (bases 1 to 720)
AUTHORS Hoischen,C., Gumpert,J., Kujau,J.M., Fritsche,C., Eiske,G.,
Fahner,B., Sieben,S. and Mueller,H.P.
TITLE Novel 1-form bacterial strains, method for producing same and the
use thereof for producing gene products
JOURNAL Patent: WO 0166776-A 15-13-SEP-2001;
INSTITUT FUER MOLEKULARE BIOLOGIE E.V. (DE)

FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/db_xref="taxon:562"
CDS 1..720
/note="LacYH1-3-Sak"
/codon_start=1
/transl_table=11
/protein_id="CAC88586.1"
/db_xref="GI:15862232"

/translation="MYLKNINFMFGLEFFEFYFIMGAYEFPPIWLDHINHSKSD
TGIIPALISLFLQPLRGLISDKIGLKKYLLITGMVLVFAFPFIIFGPILOYL
OSSSEFDKGRYKKGDDASVEEPYILMVTGVGDKRNELISPRYVEPIKGTTLTK
EKIEYVEVALADATAYKEFRVVELPSAKIEVYVDKNNKKEETKSPITEKGFVVD
LSEHIKNPQNITIVIEKK"

BASE COUNT 226 a 127 c 136 g 231 t

ORIGIN

alignment_scores: Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:
US-09-728-670-10 x AX247539 ..

Align seg 1/1 to: AX247539 from: 1 to: 720

```

1 SerSerSerPheaspLysGlyLysTyLysGlyaspaspalaSerty 17
|||||
310 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGATTA 359
17 rPhegluProthrGlyProtyrLeuMetValasnValThrGlyValasps 34
|||||
360 TTTTGAACCAACAGCCCGCTATTGTGTAATGTGACTGAGGAGTTGATG 409
34 erLysGlyasnGluLeuSerProHisTyrValGluPheProileLys 50
|||||
410 GTTAAAGAAATGATTCATTCCTCCCTGATGCGAGTTCTTATTA 459
51 ProGlyThrThrLeuThrLysGlyLysIleGlyTyTyValGluTTPAI 67
|||||
460 CCGGCGACTACACTTACAAAAGAAAATTGAATCTATGTCGAATGGCG 509
67 aleuaspalaThrAlaTyryLysGluPheargValValGluLeuaspPro 84
|||||
510 ATTAGATCGACAGCATATTAAGAGTTTAGAGTAGTGAATTAGATCCAA 559
84 erAlaLysIleGluValThrTyTyTyAspLysAsnLysLysGlyGlu 100
|||||
560 GCGCAAAAGATCGAAGTCACTTATATGATTAAGATAAGAAAAGAAAGA 609
101 ThrLysSerPheProIleThrGlyLysGlyPheValProaspLeuSe 117
|||||
610 ACGAAGTCTTCCCTATTAACAGAAAAGGTTGTGTGCCAGATTATTC 659
117 rGluHisIleLysasnProGlyPheasnLeuIleThrLysValValIleG 134
|||||
660 AGAGCATATTAATAAACCTGATTCACCTTAATTAACAAAAGTTGTATAG 709
134 LuLysLys 136
|||||
710 AAAAGAAA 717

```

seq_name: gb_pat:AX247543

seq_documentation_block:

LOCUS AX247543 876 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 19 from Patent WO0166776.
ACCESSION AX247543
VERSION AX247543.1 GI:15862236
KEYWORDS

```

SOURCE      Escherichia coli.
ORGANISM    Escherichia coli
REFERENCE   1 (bases 1 to 876)
AUTHORS     Holtschen,C., Gumpert,J., Kujau,J.M., Fritsche,C., Elske,G.,
            Fahner,L.B., Sieben,S. and Mueller,H.P.
TITLE       Novel 1-form bacterial strains, method for producing same and the
            use thereof for producing gene products
JOURNAL     Patent: WO 0166776-A 19 13-SEP-2001;
            INSTITUT FUER MOLEKULARE BIOLOGIE E.V. (DE)
FEATURES
  source     1..876
            /organism="Escherichia coli"
            /db_xref="taxon:562"
            /note="SecYH1-3-Sak"
            /codon_start=1
            /transl_table=11
            /protein_id="CAC86588.1"
            /db_xref="GI:15862237"
            /translation="MAKQPCLDPSAKGGLGELKRLFLVIGALIVRIGSFPIPIPI
            DAAVLAKLEQORGTITFEMFNESGALSRASIFALGIMPYISAIITDLYVHPTL
            AEIKKESGSRKISQYTRGTLYLAIFSGISGATLPMMPGMOGLVINGLOSSFD
            KKKYKGDADSEEPSPYLAIVNTGVQDKRRELLSPRVPEPKKGTLLTKETIYY
            VERALDATIKERFVEELDPKAKIEVTYIDKKKKETKSPITEKGFVVPDSEHIK
            NPGFNLITKVIIEKK"

BASE COUNT  259 a      168 c      207 g      242 t
ORIGIN
1  SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTyr 17
   |||||
466 TCAAGTTCATTCGACAAAGAAAATATATAAAGCCGATGACCGCAGTTA 515
   |||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
   |||||
516 TTTGAACCAACAGGCCCGTATTGATGCTAAATGTGACTGAGATTGATG 565
   |||||
34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
   |||||
566 GTTAAAGAAATGAATTCCTATCCCTCGTATGTCGAGTTTCCTAATTA 615
   |||||
51 ProGlyThrLeuThrLysGlyLysIleGluTyrTyrValGluTyrPAl 67
   |||||
616 CCTGGAGACTACCTACAAAGAAAATTAATGATACATGTCGAAATGGGC 665
   |||||
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
   |||||
666 ATTAGATGCGACACATTAATAAGATTAGAGTAGTGAATTAAGTCAAA 715
   |||||
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
   |||||
716 GCGCAAAAGATGAAAGTCATTATATGATTAAGAAATTAAGAAAAAGAA 765
   |||||
101 ThrLysSerPheProIleThrGlyLysGlyPheValValProAspLeuS 117
   |||||
766 ACGAAGTTCCTCTATACAGAAAAGGTTTGTGTCCAGATTATTC 815
   |||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
   |||||
816 AGACATATTAATAAACCTGATTCACACTTAATTAACAAAGGTTTATAG 865

```

```

134 LuLysLys 136
   |||||
866 AAAAGAAA 873
seq_name: gb_ph:P42STK
seq_documentation_block:
LOCUS      P42STK          993 bp      DNA      linear      PHG 24-APR-1996
DEFINITION Bacteriophage 42D (clone pDB17) (from Staphylococcus aureus)
KEYWORDS   M57455
VERSION    M57455.1 GI:215344
SOURCE     staphylokinase.42D (clone: pDB17) DNA.
ORGANISM   Bacteriophage 42D
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
            Lambda phage group.
REFERENCE   1 (bases 1 to 993)
AUTHORS     Behnke,D., Gerlach,D. and Kraft,R.
TITLE       Cloning, sequencing, and expression in E. coli and B. subtilis of a
            staphylokinase gene
JOURNAL     (in) Chaioupa,J. and Krumphanz,V. (Eds.);
            EXTRACELLULAR ENZYMES OF MICROORGANISMS: 29-38;
            Plenum Press, New York (1987)
FEATURES
  source     1..993
            /organism="Bacteriophage 42D"
            /specific_host="Staphylococcus aureus"
            /db_xref="taxon:10715"
            /clone="pDB17"
            /protein_id="PDB17"
            /transl_table=11
            /product="staphylokinase"
            /protein_id="AA08206.1"
            /db_xref="GI:215345"
            /translation="MLKRSLLFLTVLLLLFSFSSTINENVSASSSPDKKKKGDADY
            FEPTGPLYMNTGVQDKRRELLSPRVPEPKKGTLLTKETIYYVERALDATIKERF
            VVEELDPKAKIEVTYIDKKKKETKSPITEKGFVVPDSEHIKPNLITKVI
            EKK"
            /note="putative"
            /gene="staphylokinase"
            /note="putative"

BASE COUNT  356 a      118 c      183 g      336 t
ORIGIN
1  SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTyr 17
   |||||
395 TCAAGTTCATTCGACAAAGAAAATATATAAAGCCGATGACCGCAGTTA 444
   |||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
   |||||
445 TTTGAACCAACAGGCCCGTATTGATGCTAAATGTGACTGAGATTGATG 494
   |||||
34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50

```



```

495 GTAAAGAAATGATGCTATCCCTCGTTATGCGAGTTCTTAA 544
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPAl 67
|||||
545 CCGGAGACTTACACTTACAAAAGAAAAATTCATCTATGTCGATGGCC 594
|||||
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
595 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTTGAATGATCCAA 644
|||||
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
645 GCGCAAAAGATCGAAGTCACTTATGATGATGAAGAAAAAGAAAGA 694
|||||
101 ThrLysSerPheProIleThrGluLysGlyPheValAlaProAspLeuS 117
|||||
695 ACGAAGTCTTCCCTATACAGAAAAAGTTTGTGTCCAGATTTATAC 744
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
745 AGAGCATATTAAACCCGTGATTCACATTAATCAAGAGTTGTATAG 794
|||||
134 LuLysLys 136
|||||
795 AAAAGAAA 802

seq_name: gb_ba:SASAK42D

seq_documentation_block:
LOCUS SASAK42D 1023 bp DNA linear BCT 30-MAR-1995
DEFINITION Staphylococcus aureus phage 42D for staphylokinase.
ACCESSION X06603
VERSION X06603.1 GI:46676
KEYWORDS plasmidogen activator; sak42D gene; staphylokinase.
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
1 (bases 31 to 1023)
Behnke,D. and Gerlach,D.
Cloning and expression in Escherichia coli. Bacillus subtilis, and
Streptococcus sanguis of a gene for staphylokinase--a bacterial
plasmidogen activator
Mol. Gen. Genet. 210 (3), 528-534 (1987)
88121731
2 (bases 1 to 30)
Behnke,D.
Direct Submission
Submitted (09-JUL-1988)
Data kindly reviewed (09-JUL-1988) by Behnke D.
FEATURES
Location/Qualifiers
source
1..1023
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
224..229
/promoter
248..253
/promoter
332..336
/note="10 region"
332..336
/note="Shine-Dalgarno sequence"
344..835
CDS
1..1023
/transl_start=1
/product="sak42D staphylokinase"
/protein_id="CA29823.1"
/db_xref="GI:758275"
/db_xref="SWISS-PROT:P15240"
/transl_except="MLKRSLLFTVLLLFSSITNEVSASSFDKGYKKKGDDASY
FEPTGPYLVAVTGVCDKRNELSPRYVEPFIKGTTLTKRKIEYVYEMALDATAVKE
FRVVELDPSAKIEVYDYDKKKKEETKSFPITEKGFVVPDLSEHIKNPGFNLTKVVI
EKK"
sig_peptide
344..415
mat_peptide
416..832

```

```

BASE COUNT 365 a 123 c 190 g 345 t
ORIGIN
alignment_scores:
Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:
US-09-728-670-10 x SASAK42D ..

Align seg 1/1 to: SASAK42D from: 1 to: 1023

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTyr 17
|||||
425 TCAGATTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGCTTA 474
|||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
475 TTTTGAAACCAACAGGCGCCGTATTGATGTAAATGTGACTGAGTTGATG 524
|||||
34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|
525 GTAAAGAAATGAAATGCTATCCCTCGTATGTCGAGTTCTTAA 574
|||||
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPAl 67
|||||
575 CCGGAGACTTACACTTACAAAAGAAAAATTCATCTATGTCGATGGCC 624
|||||
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
625 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTTGAATGATCCAA 674
|||||
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
675 GCGCAAAAGATCGAAGTCACTTATGATGATGAAGAAAAAGAAAGA 724
|||||
101 ThrLysSerPheProIleThrGluLysGlyPheValAlaProAspLeuS 117
|||||
725 ACGAAGTCTTCCCTATACAGAAAAAGTTTGTGTGTCGAGATTTATC 774
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
775 AGAGCATATTAAACCCGTGATTCACATTAATCAAGAGTTGTATAG 824
|||||
134 LuLysLys 136
|||||
825 AAAAGAAA 832

seq_name: gb_pat:A17537

seq_documentation_block:
LOCUS A17537 1023 bp DNA linear PAT 19-APR-1994
DEFINITION Bacteriophage 42D gene for staphylokinase 42D.
ACCESSION A17537
VERSION A17537.1 GI:512341
KEYWORDS Bacteriophage 42D.
SOURCE Bacteriophage 42D.
ORGANISM Bacteriophage 42D.
Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
Lambda phage group.
REFERENCE
1 (bases 1 to 1023)
AUTHORS
TITLE EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES
JOURNAL Patent: WO 931209-A 9 08-JUL-1993;
FEATURES
Location/Qualifiers
1..1023
/organism="Bacteriophage 42D"
/db_xref="taxon:10715"
-35_signal
-10_signal
224..229
248..253

```

RBS 332. .336
CDS 344. .835
/codon_start=1
/product="staphylokinase 42D"
/protein_id="CA00341.1"
/db_xref="GI:512342"
/db_xref="SWISS-PROT:P15240"
/translation="MKRSLFLVLLLFSPSTINEVSASSPDKCKYKKGDADSY
FEPTGPYLMVNTGVGDKRNELSPRYVEFPKPTTLKEKLEYVEWALDAPAYKE
FRVVELDPSAKIEVTVYDKKKKEETKSPITEKGVVDLSEHIKNPGFNLTITKVI
EKK"

sig_peptide 344. .415
mat_peptide 416. .832
/product="staphylokinase 42D"
BASE COUNT 365 a 123 c 190 g 345 t
ORIGIN

alignment_scores:

Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:

US-09-728-670-10 x A17537 ..

Align seg 1/1 to: A17537 from: 1 to: 1023

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
|||||
425 TCAGGTCATTCGACAAAGAAATATATAAGCGCATGACGCGAGTTA 474
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
475 TTTTGAAACCAACAGCCCGTATTGATGGTAAATGTGACTGAGATTGATG 524
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|
525 GTAAAGAAATGAATGCTATCCCTCGTATGTCGAGTTTCCTATATAA 574
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPA 67
|||||
575 CCTGGACTACACTTACAAAGAAATTTGAAATCTATGTCGATGGGC 624
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
625 ATTAGATCGCAGCAGCATATAAAGATTAGAGTTGATTAATTAATCCAA 674
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
675 GCGCAAGATCGAAGTCACCTTATATGATTAAGAAATAGAAAAAGAAAGAA 724
725 ACAGAGTCTTCCCTATACAGAAAAAGGTTTGTGTGCCAGATTATTC 774
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
117 TGIuHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
775 AGAGCATATTAATAAACCTGATCACTTAATTAACAAGAGTTGTTATAG 824
134 LuLysLys 136
|||||
825 AAAAGAAA 832

```

seq_name: gb_pat:AR037177

seq_documentation_block:

LOCUS AR037177 1023 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 16 from patent US 5801037.
ACCESSION AR037177
VERSION AR037177.1 GI:5955033
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1023)
AUTHORS Behnke,D., Schiott,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.
TITLE Expression of signal-peptide-free staphylokinases
JOURNAL Patent: US 5801037-A 16 01-SEP-1998;
FEATURES Location/Qualifiers
source 1..1023
/organism="unknown"
BASE COUNT 365 a 123 c 190 g 345 t
ORIGIN

alignment_scores:

Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:

US-09-728-670-10 x AR037177 ..

Align seg 1/1 to: AR037177 from: 1 to: 1023

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
|||||
425 TCAGGTCATTCGACAAAGAAATATATAAGCGCATGACGCGAGTTA 474
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
475 TTTTGAAACCAACAGCCCGTATTGATGGTAAATGTGACTGAGATTGATG 524
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|
525 GTAAAGAAATGAATGCTATCCCTCGTATGTCGAGTTTCCTATATAA 574
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPA 67
|||||
575 CCTGGACTACACTTACAAAGAAATTTGAAATCTATGTCGATGGGC 624
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
625 ATTAGATCGCAGCAGCATATAAAGATTAGAGTTGATTAATTAATCCAA 674
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
675 GCGCAAGATCGAAGTCACCTTATATGATTAAGAAATAGAAAAAGAAAGAA 724
725 ACAGAGTCTTCCCTATACAGAAAAAGGTTTGTGTGCCAGATTATTC 774
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
117 TGIuHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
775 AGAGCATATTAATAAACCTGATCACTTAATTAACAAGAGTTGTTATAG 824
134 LuLysLys 136
|||||
825 AAAAGAAA 832

```

seq_name: gb_pat:AR037174

seq_documentation_block:

LOCUS AR037174 414 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 11 from patent US 5801037.
ACCESSION AR037174
VERSION AR037174.1 GI:5955030
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 414)
AUTHORS Behnke,D., Schiott,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.

TITLE Expression of signal-peptide-free staphylokinases
JOURNAL Patent: US 5801037-A 11 01-SEP-1998;
FEATURES Location/Qualifiers
source 1..414
BASE COUNT 158 a 60 c 81 g 106 t 9 others
ORIGIN

alignment_scores:
Quality: 689.00 Length: 136
Ratio: 5.142 Gaps: 0
Percent Similarity: 98.529 Percent Identity: 97.059

alignment_block:
US-09-728-670-10 x AR037174 ..

Align seg 1/1 to: AR037174 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTy 17
|||||
4 TCAGATTTCATTCGACAAAGCAAAATATATAAAAGCGATGACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAACCAACAGGCCCGTATTGTGCTAAATGTGACTGAGAGTGGAYG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
104 GTAAAGAAATGACACTTACCAAAAGAAAAATTGAATCTATGCGAATGGGC 153
51 ProGlyThrThrLeuThrLysGluLysIleGlyTyrValGluTyrAl 67
|||||
154 CCGGAGACTACACTTACCAAAAGAAAAATTGAATCTATGCGAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspPro 84
|||||
204 ATTAGATGCGACAGCATATAAGAGTTAGAGTAGATTGAATTAGATCCAA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysLysGlu 100
|||||
254 GCCCAAAAGATCGAAGCTATTATGATACAGATTAAGAAAAAGAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValAlaProAspLeu 117
|||||
304 ACGAAGCTTTCCCTATTAACAGAAAAAGTTTGTGTCCTCCAGATTATTC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValAlaIle 134
|||||
354 AGAGCATATTATAAAACCTGATTCAACTTAATACCAAGGTGTATATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411

```

seq_name: gb_pat:AR037175

seq_documentation_block:
LOCUS AR037175 414 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 13 from patent US 5801037.
ACCESSION AR037175
VERSION AR037175.1 GI:5955031
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 414)
AUTHORS Behnke,D., Schlotz,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.
TITLE Expression of signal-peptide-free staphylokinases
JOURNAL Patent: US 5801037-A 11 01-SEP-1998;
FEATURES Location/Qualifiers
source 1..414
/organism="unknown"

BASE COUNT 159 a 60 c 80 g 106 t 9 others
ORIGIN

alignment_scores:
Quality: 687.00 Length: 136
Ratio: 5.127 Gaps: 0
Percent Similarity: 98.529 Percent Identity: 96.324

alignment_block:
US-09-728-670-10 x AR037175 ..

Align seg 1/1 to: AR037175 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTy 17
|||||
4 TCAGATTTCATTCGACAAAGCAAAATATATAAAAGCGATGACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAACCAACAGGCCCGTATTGTGCTAAGTAAATGTGACTGAGAGTGGAYG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
104 GTAAAGAAATGACACTTACCAAAAGAAAAATTGAATCTATGCGAATGGGC 153
51 ProGlyThrThrLeuThrLysGluLysIleGlyTyrValGluTyrAl 67
|||||
154 CCGGAGACTACACTTACCAAAAGAAAAATTGAATCTATGCGAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspPro 84
|||||
204 ATTAGATGCGACAGCATATAAGAGTTAGAGTAGATTGAATTAGATCCAA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysLysGlu 100
|||||
254 CGCAGAAAGATCGAAGCTACTTATATGATTAAGATAAGAAAAAGAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValAlaProAspLeu 117
|||||
304 ACGAAGCTTTCCCTATTAACAGAAAAAGTTTGTGTCCTCCAGATTATTC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValAlaIle 134
|||||
354 AGAGCATATTATAAAACCTGATTCAACTTAATACMAAGCTGTATATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411

```

seq_name: gb_pat:A17534

seq_documentation_block:
LOCUS A17534 414 bp DNA linear PAT 19-APR-1994
DEFINITION Bacteriophage 42D DNA for staphylokinase SARW26C.
ACCESSION A17534
VERSION A17534.1 GI:512337
KEYWORDS
SOURCE Bacteriophage 42D.
ORGANISM Bacteriophage 42D.
REFERENCE 1 (bases 1 to 414)
AUTHORS
TITLE Lambda phage group.
JOURNAL Viruses; dsDNA viruses, no RNA stage; Caudovirales; siphoviridae;
FEATURES Patent: WO 9313209-A 6 08-JUL-1993;
source 1..414
Location/Qualifiers
CDS 1..414
/db_xref="taxon:10715"
/codon_start=1
/transl_table=11

/product="staphylokinase SAKM26C"
/protein_id="CAA01339.1"
/db_xref="GI:512338"
/translation="MSSSPDKGKYYKKGDASYEPFPGYLVAVNTGVXGKRNXLSPR
YVEPPIKGGTTLTKREKIEYVEMALDATALYKKEFVVELDPSAKIEVITYYDKNNKKKEET
KSPFITEKGFFVDPDSEHINKNPFNLTKVIEKK"

BASE COUNT 158 a 60 c 81 g 106 t 9 others
ORIGIN

alignment_scores:
Quality: 671.00 Length: 136
Ratio: 5.122 Gaps: 0
Percent Similarity: 96.324 Percent Identity: 94.853

alignment_block:

US-09-728-670-10 x A17534 ..

Align seg 1/1 to: A17534 from: 1 to: 414

```
1 SerSerSerPheAspLysGlyLysTyrlLysGlyAspAspAlaSerTy 17
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 53
|||||
17 rPheGluProThrGlyProTyrlLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAACCAACAGCGCCCGTATTTGTNGTAAATGTGACTGCGAGTNGANG 103
|||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrlValGluPheProIleLys 50
::|||
104 GTAAAGAAATGANCNTNNTNNTCCCGCNGTATGTCGAGTTCCATTTAA 153
|||||
51 ProGlyThrThrlLeuThrlLysGlyLysIleGluTyrlValGluTrpAl 67
|||||
154 CCGTGGACTACACTTACAAAAGAAAATTTGAATGATGTCGAATGGGC 203
|||||
67 AleuAspAlaThrAlaTyrlLysGlyLysPheArgValGluLeuAspPro 84
|||||
204 ATTAGATGCGACACATTTAAAGAGTTAGAGTTGAATTAAGATCCAA 253
|||||
84 erAlaLysIleGluValThrTyrlTyrlAspLysAsnLysLysGlyLys 100
|||||
254 GCGCAAGATGCAAGTCACTATATATGATTAAGATTAAGAAAAGAAAGA 303
|||||
101 ThrLysSerPheProIleThrGlyLysGlyPheValAlaProAspLeu 117
|||||
304 ACGAAGCTCTTCCCTATTAACAGAAAAGGTTTGTGTCCACAGATTATC 353
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrlLysValAlaIleG 134
|||||
354 AGACCATATTAAACCCGTGATTCAACTTAATTATCAACAGGTTGTTATAG 403
|||||
134 LysLysLys 136
|||||
404 AAAAGAAA 411
```

seq_name: gb_pat:A17535

seq_documentation_block:

LOCUS A17535 414 bp DNA linear PAT 19-APR-1994
DEFINITION Bacteriophage 42D DNA for staphylokinase SAKM26L.
ACCESSION A17535
VERSION A17535.1 GI:512339
KEYWORDS
SOURCE Bacteriophage 42D.
ORGANISM Bacteriophage 42D
VIRUSES: dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
Lambdaphage group.
REFERENCE 1 (bases 1 to 414)
AUTHORS
TITLE EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES
JOURNAL Patent: WO 9313209-A 7 08-JUL-1993;
FEATURES Location/Qualifiers

source 1. 414
/organism="Bacteriophage 42D"
/db_xref="taxon:10715"
1. 414
/codon_start=1
/transl_table=1
/product="staphylokinase SAKM26L"
/protein_id="CAA01340.1"
/db_xref="GI:512340"
/translation="MSSSPDKGKYYKKGDASYEPFPGYLVAVNTGVXGKRNXLSPR
YVEPPIKGGTTLTKREKIEYVEMALDATALYKKEFVVELDPSAKIEVITYYDKNNKKKEET
KSPFITEKGFFVDPDSEHINKNPFNLTKVIEKK"

BASE COUNT 159 a 60 c 81 g 105 t 9 others
ORIGIN

alignment_scores:
Quality: 666.00 Length: 136
Ratio: 5.123 Gaps: 0
Percent Similarity: 95.588 Percent Identity: 94.118

alignment_block:

US-09-728-670-10 x A17535 ..

Align seg 1/1 to: A17535 from: 1 to: 414

```
1 SerSerSerPheAspLysGlyLysTyrlLysGlyAspAspAlaSerTy 17
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 53
|||||
17 rPheGluProThrGlyProTyrlLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAACCAACAGCGCCCGTATTTGTGACGTAATGTGACTGCGAGTNGANG 103
|||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrlValGluPheProIleLys 50
::|||
104 GTAAAGAAATGANCNTNNTNNTCCCGCNGTATGTCGAGTTCCATTTAA 153
|||||
51 ProGlyThrThrlLeuThrlLysGlyLysIleGluTyrlValGluTrpAl 67
|||||
154 CCGTGGACTACACTTACAAAAGAAAATTTGAATGATGTCGAATGGGC 203
|||||
67 AleuAspAlaThrAlaTyrlLysGlyLysPheArgValGluLeuAspPro 84
|||||
204 ATTAGATGCGACACATTTAAAGAGTTAGAGTTGAATTAAGATCCAA 253
|||||
84 erAlaLysIleGluValThrTyrlTyrlAspLysAsnLysLysGlyLys 100
|||||
254 GCGCAAGATGCAAGTCACTATATATGATTAAGATTAAGAAAAGAAAGA 303
|||||
101 ThrLysSerPheProIleThrGlyLysGlyPheValAlaProAspLeu 117
|||||
304 ACGAAGCTCTTCCCTATTAACAGAAAAGGTTTGTGTCCACAGATTATC 353
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrlLysValAlaIleG 134
|||||
354 AGACCATATTAAACCCGTGATTCAACTTAATTATCAACAGGTTGTTATAG 403
|||||
134 LysLysLys 136
|||||
404 AAAAGAAA 411
```

seq_name: gb_pat:A17532

seq_documentation_block:

LOCUS A17532 384 bp DNA linear PAT 06-JUL-1995
DEFINITION Bacteriophage 42D DNA for staphylokinase SAK-delta-n10 (short).
ACCESSION A17532
VERSION A17532.1 GI:512333
KEYWORDS
SOURCE Bacteriophage 42D.
ORGANISM Bacteriophage 42D
VIRUSES: dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;

```

REFERENCE      1 (bases 1 to 384)
AUTHORS
TITLE          EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES
JOURNAL        Patent: WO 9313209-A 4 08-JUL-1993;
FEATURES
SOURCE
CDS
  /organism="Bacteriophage 42D"
  /db_xref="taxon:10715"
  <1..384
  /note="short form"
  /codon_start=1
  /transl_table=11
  /product="staphylokinase SAK-delta-N10"
  /protein_id="CAA01337.1"
  /db_xref="GI:512334"
  /translation="MKGDASYPEPTGYLVNVTGVDGRNELSPRYVEPIKPGT
  TLTKRIEYVEVALDATALKEPRVVELDPSAKIEVYYDKNKKKEETTSFPTENG
  VVPDLSEHINPQFNLTITKVIVIEK"
BASE COUNT    146 a      55 c      78 g      104 t      1 others
ORIGIN
alignment_scores:
  Quality: 643.00      Length: 126
  Ratio: 5.144        Gaps: 0
  Percent Similarity: 99.206      Percent Identity: 97.619

alignment_block:
US-09-728-670-10 x A17532  ..

Align seg 1/1 to: A17532 from: 1 to: 384

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27
|||||
4 AAAGCGATGACGCGAGTATTGTAACCAACAGCCCGCTATTATGCT 53
27 IAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisT 44
|||||
54 AAATGTGACTGAGTGTGATGTAAGAAAGAAATGATGCTATCCCTCGTT 103
44 yValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60
|||||
104 ATGCGAGTTTCCCTATTAACCTGCGACTACACTTACAAAAGAAAAATT 153
61 GluTyrTyrValGluTrrPalaLeuAspAlaThrAlaTyrLysGluPheAr 77
|||||
154 GAATACCTATGTCGANTGGCATTAGATCGACAGCATATAAAGACTTTAG 203
77 yValValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspL 94
|||||
204 AGTAGTTGAATTAGATCCAAAGCGCAAGATCGAAGTCACTTATATGATA 253
94 yAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
|||||
254 AGAATTAAGAAAAAGAAAGAAAGAAAGTCTTCCCTATTAACAGAAAAAGGT 303
111 PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnLe 127
|||||
304 TTGTGTGTCGCCAGATTATACAGACATATAAACCCTGGATTCAACTT 353
127 uIleThrLysValValIleGluLysLys 136
|||||
354 AATTACNAAGGTTGTATAGAAAAAGAAA 381

seq_name: gb_pat:AR037172
seq_documentation_block:
LOCUS          AR037172      384 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION     Sequence 7 from patent US 5801037.
ACCESSION      AR037172
VERSION        AR037172.1  GI:5955028
KEYWORDS

```

```

SOURCE          Unknown.
ORGANISM        Unknown.
REFERENCE       Unclassified.
AUTHORS         1 (bases 1 to 384)
TITLE          Behke,D., Schloett,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.
JOURNAL        Expression of signal-peptide-free staphylokinases
FEATURES
SOURCE
  /organism="unknown"
BASE COUNT      146 a      55 c      78 g      104 t      1 others
ORIGIN
alignment_scores:
  Quality: 643.00      Length: 126
  Ratio: 5.144        Gaps: 0
  Percent Similarity: 99.206      Percent Identity: 97.619

alignment_block:
US-09-728-670-10 x AR037172  ..

Align seg 1/1 to: AR037172 from: 1 to: 384

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27
|||||
4 AAAGCGATGACGCGAGTATTGTAACCAACAGCCCGCTATTATGCT 53
27 IAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisT 44
|||||
54 AAATGTGACTGAGTGTGATGTAAGAAAGAAATGATGCTATCCCTCGTT 103
44 yValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60
|||||
104 ATGCGAGTTTCCCTATTAACCTGCGACTACACTTACAAAAGAAAAATT 153
61 GluTyrTyrValGluTrrPalaLeuAspAlaThrAlaTyrLysGluPheAr 77
|||||
154 GAATACCTATGTCGANTGGCATTAGATCGACAGCATATAAAGACTTTAG 203
77 yValValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspL 94
|||||
204 AGTAGTTGAATTAGATCCAAAGCGCAAGATCGAAGTCACTTATATGATA 253
94 yAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
|||||
254 AGAATTAAGAAAAAGAAAGAAAGAAAGTCTTCCCTATTAACAGAAAAAGGT 303
111 PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnLe 127
|||||
304 TTGTGTGTCGCCAGATTATACAGACATATAAACCCTGGATTCAACTT 353
127 uIleThrLysValValIleGluLysLys 136
|||||
354 AATTACNAAGGTTGTATAGAAAAAGAAA 381

seq_name: gb_pat:A17533
seq_documentation_block:
LOCUS          A17533      372 bp      DNA      linear      PAT 19-APR-1994
DEFINITION     Bacteriophage 42D DNA for staphylokinase SAK-delta-N14 (short).
ACCESSION      A17533
VERSION        A17533.1  GI:512335
KEYWORDS
SOURCE
  Bacteriophage 42D.
ORGANISM        Bacteriophage 42D.
  Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
  Lambda phage group.
REFERENCE       1 (bases 1 to 372)
AUTHORS
TITLE          EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES
JOURNAL        Patent: WO 9313209-A 5 08-JUL-1993;
FEATURES
  Location/Qualifiers

```

source 1. 372
/organism="Bacteriophage 420"
/db_xref="taxon:10715"
CDS 1. 372
/note="short form"
/codon_start=1
/product="staphylokinase Sak-delta-N14"
/protein_id="CA01338.1"
/db_xref="GI:512336"
/translation="MASYFEPSTGYLVNVTGVNDRNELSPRYVEPIKPGTTLTK
EKIEYVEMADATAYKKEFRVEVDPSAKLEVITYYDKNKKKEETKSPITEKGVVD
LSEHKNGEFLTKVIEKK"
BASE COUNT 141 a 53 c 74 g 103 t 1 others
ORIGIN

alignment_scores:
Quality: 620.00 Length: 122
Ratio: 5.124 Gaps: 0
Percent similarity: 99.180 Percent identity: 97.541

alignment_block:
US-09-728-670-10 x A17533 ..
Align seg 1/1 to: A17533 from: 1 to: 372

15 ALaSerTyRPhEgLuPProThrGlyProTyRLeuMeTValAsnValThrG1 31
|||||
4 GCGAGTTATTTGAACCAACAGCCCGTATTTGATGTAAATGTGACTGG 53
31 yValAspSerLySgLYAsnGluLeuLeuSerProHisTyRValGluPheP 48
|||||
54 AGTTGATGTAAAGAAATGAAATTCCTATCCCTCGTTATGTCGAGTTTC 103
48 roLleYsPProGlyThrThrLeuThrLySgLuLySgLyLeuTyRVal 64
|||||
104 CTATTAAACCTGGGACTACCTTACAAAGAAAAATTGCAATCTATGTC 153
65 GluTrpAlaLeuAspAlaThrAlaTyRlySgLuPheArgValValGluLe 81
|||||
154 GAATGGCATTAGATGGACAGCATATAAGATTAGACTAGTGAAT 203
81 uAspProSerAlaLySgLyLeuValThrTyRAspLySasnlYsLyS 98
|||||
204 AGATTCACAGCGCAAGATCGAAGTCACTTATGTATGAAGAAATMAAGAAA 253
98 ySgLuGluThrLySerPheProIleThrGluLySgLyPheValValPro 114
|||||
254 AAGAGAAACGAAGCTCTTCCCTATACAGAAAAAGCTTTTGTGTC 303
115 AspLeuSerGluHisIleLySasnProGlyPheAsnLeuIleThrLyS 131
|||||
304 GATTATACAGACATATTAAACCCGTGATCACTTAATATTCNAGGT 353
131 lValIleGluLyS 136
|||||
354 TGTATAGAAAAAGAAA 369

seq_name: gb_pat:AR037173

seq_documentation_block:
LOCUS AR037173 372 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 9 from patent US 5801037.
ACCESSION AR037173
VERSION AR037173.1 GI:5955029
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 372)
AUTHORS Behnke,D., Schloft,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.
TITLE Expression of signal-peptide-free staphylokinases

JOURNAL Patent: US 5801037-A 9 01-SEP-1998;
FEATURES location/Qualifiers
source 1. 372
BASE COUNT 141 a 53 c 74 g 103 t 1 others
ORIGIN

alignment_scores:
Quality: 620.00 Length: 122
Ratio: 5.124 Gaps: 0
Percent similarity: 99.180 Percent identity: 97.541

alignment_block:
US-09-728-670-10 x AR037173 ..
Align seg 1/1 to: AR037173 from: 1 to: 372

15 ALaSerTyRPhEgLuPProThrGlyProTyRLeuMeTValAsnValThrG1 31
|||||
4 GCGAGTTATTTGAACCAACAGCCCGTATTTGATGTAAATGTGACTGG 53
31 yValAspSerLySgLYAsnGluLeuLeuSerProHisTyRValGluPheP 48
|||||
54 AGTTGATGTAAAGAAATGAAATTCCTATCCCTCGTTATGTCGAGTTTC 103
48 roLleYsPProGlyThrThrLeuThrLySgLuLySgLyLeuTyRVal 64
|||||
104 CTATTAAACCTGGGACTACCTTACAAAGAAAAATTGCAATCTATGTC 153
65 GluTrpAlaLeuAspAlaThrAlaTyRlySgLuPheArgValValGluLe 81
|||||
154 GAATGGCATTAGATGGACAGCATATAAGATTAGACTAGTGAAT 203
81 uAspProSerAlaLySgLyLeuValThrTyRAspLySasnlYsLyS 98
|||||
204 AGATTCACAGCGCAAGATCGAAGTCACTTATGTATGAAGAAATMAAGAAA 253
98 ySgLuGluThrLySerPheProIleThrGluLySgLyPheValValPro 114
|||||
254 AAGAGAAACGAAGCTCTTCCCTATACAGAAAAAGCTTTTGTGTC 303
115 AspLeuSerGluHisIleLySasnProGlyPheAsnLeuIleThrLyS 131
|||||
304 GATTATACAGACATATTAAACCCGTGATCACTTAATATTCNAGGT 353
131 lValIleGluLyS 136
|||||
354 TGTATAGAAAAAGAAA 369

seq_name: gb_pat:E00876

seq_documentation_block:
LOCUS E00876 634 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA fragment comprising a promoter of sak gene and the region
coding the signal peptide.
ACCESSION E00876
VERSION E00876.1 GI:2169137
KEYWORDS JP 1986135590-A/1.
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE 1 (bases 1 to 634)
AUTHORS Sako,T..
TITLE COMPOSITE PLASMID
JOURNAL Patent: JP 1986135590-A 1 23-JUN-1986;
YAKULT HONSHA CO LTD
OS Staphylococcus aureus
PN JP 1986135590-A/1
PD 23-JUN-1986
PF 05-DEC-1984 JP 1984257258
PI SAKO TOMOYUKI

REFERENCE 1 (bases 1 to 492)
 AUTHORS Yoshimoto, M., Hanada, K. and Tamaki, S.
 TITLE TRYPSTATIN STRUCTURAL GENE AND USE THEREOF
 JOURNAL Patent: JP 1992112791-A 3 14-APR-1992;
 TAIHO PHARMACEUT CO LTD

COMMENT
 OS Artificial gene
 OC Artificial sequence; Genes.
 PN JP 1992112791-A/3
 PD 14-APR-1992
 PI YOSHIMOTO MAKOTO, HANADA KAZUNORI, TAMAKI SHIGEO PC
 C12N15/15, C12N1/21, C12N9/99, C12N15/62, (C12N1/21, C12R1:19), PC
 (C12N1/21,
 PC C12R1:865), (C12N9/99, C12R1:19), (C12N9/99, C12R1:865); CC
 strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 FH Key Location/Qualifiers
 FH 5'UTR 1..51
 FT misc_signal 40..44
 FT sig_peptide 52..132 /note='shine Dalgarno sequence' FT
 FT CDS 133..300 /product='staphylokinase signal peptide' FT
 FT CDS 301..486 /product='staphylokinase'
 FT CDS 487..492 /product='trypstatin'
 FT 3'UTR Location/Qualifiers
 source 1..492
 /organism='synthetic construct'
 /db_xref='taxon:32630'

BASE COUNT 151 a 90 c 107 g 144 t
 ORIGIN

alignment_scores:
 Quality: 277.00 Length: 52
 Ratio: 5.327 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.077

alignment_block:
 US-09-728-670-10 x E03544 ..

Align seg 1/1 to: E03544 from: 1 to: 492

```

1 SerSerSerPheAspIysGlyIySTyrlYsGlyAspAlaSerTy 17
|||||
133 TCAAGTTTCATTGCGACAAAGAAATATATAAAAGCGATGACCGGATTA 182
17 rPheGluPrOthrGlyProTyrlLeuMetValAsnValThrGlyValAsp 34
|||||
183 TTTTGAACCAACAGCCCGGTATTGATGTAATGTGACTGGAGTTGATG 232
34 eRlysgIyAsnGluLeuLeuSerProHisTyrlValGluPheProIleYs 50
:::
233 GTAAGGAATGAAATTCCTATCCCTCATATATGCGAGTTTCTATTAA 282
51 ProGly 52
|||||
283 CCTGGG 288

```

seq_name: gb_pat:E03543
 LOCUS E03543 494 bp DNA linear PAT 29-SEP-1997
 DEFINITION DNA sequence coding for staphylokinase signal peptide,
 staphylokinase N-terminus peptide and trypstatin.
 ACCESSION E03543
 VERSION E03543.1 GI:2171759

KEYWORDS JP 1992112791-A/2.
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 494)
 AUTHORS Yoshimoto, M., Hanada, K. and Tamaki, S.
 TITLE TRYPSTATIN STRUCTURAL GENE AND USE THEREOF
 JOURNAL Patent: JP 1992112791-A 2 14-APR-1992;
 TAIHO PHARMACEUT CO LTD

COMMENT
 OS Artificial gene
 OC Artificial sequence; Genes.
 PN JP 1992112791-A/2
 PD 14-APR-1992
 PI YOSHIMOTO MAKOTO, HANADA KAZUNORI, TAMAKI SHIGEO PC
 C12N15/15, C12N1/21, C12N9/99, C12N15/62, (C12N1/21, C12R1:19), PC
 (C12N1/21,
 PC C12R1:865), (C12N9/99, C12R1:19), (C12N9/99, C12R1:865); CC
 strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 FH Key Location/Qualifiers
 FH 5'UTR 1..59
 FT misc_signal 48..52
 FT sig_peptide 60..140 /note='shine Dalgarno sequence' FT
 FT met_peptide 141..299 /product='staphylokinase signal peptide' FT
 FT CDS 300..488 /product='staphylokinase'
 FT CDS 489..494 /product='trypstatin'
 FT 3'UTR Location/Qualifiers
 source 1..494
 /organism='synthetic construct'
 /db_xref='taxon:32630'

BASE COUNT 152 a 91 c 106 g 145 t
 ORIGIN

alignment_scores:
 Quality: 277.00 Length: 52
 Ratio: 5.327 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.077

alignment_block:
 US-09-728-670-10 x E03543 ..

Align seg 1/1 to: E03543 from: 1 to: 494

```

1 SerSerSerPheAspIysGlyIySTyrlYsGlyAspAlaSerTy 17
|||||
141 TCAAGTTTCATTGCGACAAAGAAATATATAAAAGCGATGACCGGATTA 190
17 rPheGluPrOthrGlyProTyrlLeuMetValAsnValThrGlyValAsp 34
|||||
191 TTTTGAACCAACAGCCCGGTATTGATGTAATGTGACTGGAGTTGATG 240
34 eRlysgIyAsnGluLeuLeuSerProHisTyrlValGluPheProIleYs 50
:::
241 GTAAGGAATGAAATTCCTATCCCTCATATATGCGAGTTTCTATTAA 290
51 ProGly 52
|||||
291 CCTGGG 296

```

seq_name: gb_pat:SUB314852
 LOCUS SUB314852 2058 bp DNA linear BCT 21-DEC-2001
 DEFINITION Streptococcus uberis paub gene, ORF2 and ORF3.

ACCESSION	AJ314852
VERSION	AJ314852.1
KEYWORDS	ORF2; ORF3; paub gene; plasmidogen activator.
SOURCE	Streptococcus uberis.
ORGANISM	Streptococcus uberis Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
REFERENCE	1 (sites)
AUTHORS	Ward P.N. and Leigh J.A.
TITLE	Characterization of Paub, a Novel Broad-Spectrum Plasmidogen Activator from Streptococcus uberis
JOURNAL	J. Bacteriol. 184 (1), 119-125 (2002)
PUBMED	11741851
REFERENCE	2 (bases 1 to 2058)
AUTHORS	Ward, P.N.
TITLE	Direct Submission
JOURNAL	Submitted (08-JUN-2001) Ward P.N., Environmental Microbiology, Institute for Animal Health, Compton Laboratory, Compton, nr Newbury, Berkshire, RG20 7NN, UNITED KINGDOM
FEATURES	location/Qualifiers
source	1..2058
	/organism="Streptococcus uberis"
	/strain="SK880"
	/db_xref="taxon:1349"
	/country="Denmark"
	109..1242
gene	/gene="paub"
	109..183
slg_peptide	/gene="paub"
	109..1242
CDS	/gene="paub"
	/function="bovine plasmidogen activation"
	/codon_start=1
	/trnasl_table=1
	/product="plasmidogen activator"
	/protein_id="CAC85651.1"
	/db_xref="GI:16215457"
	/translation="MKKLYVLTITGFGLLGIPYKVAITSKRVNYOEPQIMLTIN ITGSKREGQLLDSPYEQFRTIAGEVINKKELDLQVLVIDSASNOFEVADPEPSK VEMSYFDRHLLDITGERGFYDPDYSIEKRPSFLTGPVIIDKSPTEKTKTIDSTND VILHNSYFPMKRNESNMNLPVTTQLSSFSKTAQSGIOSIEELYQAAQSIFEGTKEY HEGGLQIKLMTSISENEKAVSYQFONGKPFNVYTIOLREFSKMDYSOAIRDDIIE NYVTSKNGDVPYQIOLITLDITNEETSIRTLIDPKENTIEIKHYISEADLTREI DAKGDHYMPSQNNKQYSSHHFQANYSKNSDLFYNNRSIQ"
	184..1239
mat_peptide	/gene="paub"
	/evidence="experimental"
	/product="plasmidogen activator"
	/complement(1295..1519)
CDS	/function="unknown"
	/note="ORF2"
	/codon_start=1
	/trnasl_table=1
	/product="hypothetical protein"
	/protein_id="CAC85652.1"
	/db_xref="GI:16215458"
	/translation="MAYIGLAPFLISFLYSLLYCLGLSPFYIFFLAAVHAGSGIGD FYIYLLVFKRKQRILEDLTLGLTIYQK"
	/complement(1536..1852)
CDS	/function="unknown"
	/note="ORF3"
	/codon_start=1
	/trnasl_table=1
	/product="hypothetical protein"
	/protein_id="CAC85653.1"
	/db_xref="GI:16215459"
	/translation="MKIIDKVRLOGNKKLLLLNLFSLPLTYLFFLLFTQAIIFKPL INKTSITSGIVMGVMALVIYHIELVHGLFYKLFKPKRKIRIKIGRLKSMIG"
BASE COUNT	762 a 362 c 319 g 615 t
ORIGIN	
alignment_scores:	

[illegible]

FEATURES Street, New Orleans, LA 70119, USA
Location/Qualifiers

source 1..1248
/organism="Streptococcus dysgalactiae subsp. equisimilis"
/strain="87-542-W"
/db_xref="taxon:119602"

gene 1..>1248
/gene="skc"
1..>1248
/gene="skc"
/function="plasmaenogen activator"

CDS
/codon_start=1
/transl_table=11
/product="Streptokinase precursor"
/protein_id="AAC83933.1"
/db_xref="GI:3983433"

BASE COUNT 459 a 247 c 201 g 341 t
ORIGIN

alignment_scores:
Quality: 138.50 Length: 135
Ratio: 1.710 Gaps: 4
Percent Similarity: 60.000 Percent Identity: 28.148

alignment_block:
US-09-728-670-10 x AF104301 ..

Align seg 1/1 to: AF104301 from: 1 to: 1248

```

7  Gilystfyrlyslsglyaspaspala.....SerTyrrh 18
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
139 GGAACCTTACCAACCAACGATGATGATTAATTCACAAAGATATTCATCA 188
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18  egluPrtHrGlyProTyrLeuMetValAsnValThrglyValAspSerL 35
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 AGACACCACTGGCTATATCTTATGATTAATGTAACGGAGTCGATAGTC 238
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35  ysglyAsnGluLeuLeuSerProHisTyrValGluPheProIleLysPro 51
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
239 GTCACACACACCTGTTTACCTCAATATCAAGATATAAAGCGACTAT 288
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52  GlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpAla 68
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
289 AATTAACCTTATACAAAGATGATGATTTGAGTTCCTTCAAAAACGTCAT 338
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68  uAspAlaThrAlaTyrLysGluPheArgValAlaGluLeuAspProSerA 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
339 TGAATGCT...TCAATCGAGAATTTTAAAGTTATGATTTCATGACGATA 365
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85  lAlslsIleGluValThrTyrTyrAsp..... 93
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
386 AGCTGTAGTAACGCTACTCAAGAGATGAGTCCCAATACTTTTCA 435
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
94  .....LysAsnLysLysLysGluGluThrLysSerPheProIleTh 107
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 ACAGTACTGTAACCAACGATGATGATGATGATGATGATGATGATGATGAT 485
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
107 rGluLysGlyPheValAlaPro.....AspLeuSerGluHis 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
486 CGGTAAGTACTTGTCAACCTGTTAGTAAAGAACCTTAAAGAACGATGA 535
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 lLys 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
536 TTAAG 540

```

seq_name: gb_pat:AX073719

seq_documentation_block:

LOCUS AX073719 66 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 12 from Patent WO0104287.
ACCESSION AX073719
VERSION AX073719.1 GI:12710137
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 66)
AUTHORS Halkier,T., Pedersen,A.H. and Okkels,J.S.
TITLE A method for preparing modified polypeptides
JOURNAL Patent: WO 0104287-A 12 18-JAN-2001;
Maxygen Aps (DK)

FEATURES
Location/Qualifiers

source

1..66
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer"

BASE COUNT 18 a 10 c 9 g 29 t
ORIGIN

alignment_scores:
Quality: 93.00 Length: 18
Ratio: 5.167 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-728-670-10 x AX073719/rev ..

Align seg 1/1 to reverse of: AX073719 from: 1 to: 66

```

119 HisIleLysAsnProGlyPheAsnLeuThrLysValAlaIleGluLys 135
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
65 CATATTAAACCTGATTCACCTTAATTAACAAAGTTGTTATGAGAAA 16
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 sLys 136
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15 GAAA 12

```

seq_name: gb_pat:AX073720

seq_documentation_block:

LOCUS AX073720 69 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 13 from Patent WO0104287.
ACCESSION AX073720
VERSION AX073720.1 GI:12710138
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 69)
AUTHORS Halkier,T., Pedersen,A.H. and Okkels,J.S.
TITLE A method for preparing modified polypeptides
JOURNAL Patent: WO 0104287-A 13 18-JAN-2001;
Maxygen Aps (DK)

FEATURES
Location/Qualifiers

source

1..69
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer"

BASE COUNT 30 a 13 c 10 g 16 t
ORIGIN

alignment_scores:
Quality: 92.00 Length: 27
Ratio: 4.000 Gaps: 2
Percent Similarity: 85.185 Percent Identity: 81.481


```

seq_name: gb_ba:AB007638
seq_documentation_block:
LOCUS      AB007638                16585 bp    DNA             linear    BCT-13-FEB-1999
DEFINITION Bacillus subtilis genomic DNA containing guta to cota region, 48
ACCESSION  AB007638
VERSION    AB007638.1
KEYWORDS   GI:2522006
SOURCE     Bacillus subtilis (strain:Marburg 168) DNA.
            Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE  1 (sites)
            Kasahara, Y., Nakai, S., Ogasawara, N., Yata, K. and Sadate, Y.
            Sequence analysis of the groESL-cota region of the Bacillus
            subtilis genome, containing the restriction/modification system
            genes
JOURNAL    DNA Res. (1997) In press
REFERENCE  2 (bases 1 to 16585)
AUTHORS    Sadate, Y.
TITLE      Direct Submission
SUBMITTED  (03-Oct-1997) Yoshito Sadate, National Institute of
            Genetics, Radioisotope Center; Yata 1111, Mishima, Shizuoka 411,
            Japan (E-mail:ysadate@lab.nig.ac.jp, Tel:81-0559-81-6870,
            Fax:81-0559-81-6870)
FEATURES   Location/Qualifiers
            1..16585
                /organism="Bacillus subtilis"
                /strain="Marburg 168"
                /db_xref="taxon:1423"
                /map="48 degree"
            1..19
                /gene="gutaB"
                /BC_number="1.1.1.14"
                /note="L-Idiotol 2-dehydrogenase"
                /codon_start=2
                /transl_table=11
                /product="sorbitol dehydrogenase"
                /protein_id="BAA22869.1"
                /db_xref="GI:4432758"
                /translation="VYPMR"
            82..1484
                /gene="gutaA"
                /BC_number="82..88"
                /gene="gutaA"
                /BC_number="93..1484"
                /gene="putaA"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="sugar transport protein"
                /protein_id="BAA22759.1"
                /db_xref="GI:2522008"
                /translation="MIGTEKVTALAKKHKSSGTGLSWPERIGYGFQDMSYNIPOFV
            MAYLLFYVNDVGIOPAVATLFIYRVLDALFDPIWGLIDKTNRMKARKARYLLMY
            APPALFTLTCTTTPHGETGMNVAIVYITLLGHSFSQOTIPVNSLTGRMNSVER
            TVLTTRMTLVYIGILSTSCATPLATAGGSDQAGFQMTALITAAVSVLNFSEF
            TYERIOPKRRKOGIKITLSYLFKNKPLMISSEFLAIGFINILSTMVYFTYV
            NKEEFVNGTVLEFGAALISNLFIPFSEKMGKQVMTITAAISLISYGLHYYTSS
            ILFIETWLFASGEFTTPLNTLAWGVADCVDAEMKGTIRADGVVTSMSKFNKLGVA
            LAGSPATVYIGAGYVANTDQTVASLNAIKNNALIPGFILSLIILAFYPLEKRY
            KIIISLEQRSAR"
            1563..2541
                /gene="fruc"
                /BC_number="1563..1574"
                /gene="fruc"

```

```

CDS
            1579..2541
                /gene="fruc"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="sugar transport protein"
                /protein_id="BAA22760.1"
                /db_xref="GI:2522009"
                /translation="MRSHSVYICIGELLIDFECTDVVDLMEGROFLKSAGAPANYSA
            AIKLGDAFSGKVGDPDFYFLKRTLDVAHYDTSMLVWDEKAPPTLFLVSLKNGE
            RDEVFVNGADALEFTLEDIOEKLENAKIIHFSATALLSDPFSATLRIMSKONGE
            FTSFDPNREDMLMRCGRVSEFVSAAKKAIVSDPVKVSDELEITISGVKHEKVALIH
            EIGANIVATVLGKSTLNSGKREITPSTIPVSTDSGAGDAFVGAAALYQLANTDQI
            QSVDADEFVLRLEIVAPANKVAGLVCTKIGALDPLSLDEIGVSL"
            2722..2727
                /gene="yjdJ"
                /BC_number="2722..3420"
                /gene="yjdJ"
                /BC_number="2737..3420"
                /gene="yjdJ"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="membrane protein"
                /protein_id="BAA22761.1"
                /db_xref="GI:2522010"
                /translation="MSIIGFKDIMSANINALLDKAENPEKVVQYLRNNSDLAKV
            AETAAVVAEQRRAKREYHCGADMEKESYAMALQAGNSDARKFLERTSLESKLS
            ELQANQIATNAQOMKMHDKVSDIGELEAKRMKIKMVAVAKTOEPMNKIGASVS
            SRSQMSAFRCRMEDKYNKALDQANMAELNSARQDMADLSAKYIDGSSQVDELLA
            LKAKMLDK"
            join(3472..3479,3486..4511)
                /gene="yjdG"
                /BC_number="3472..3479"
                /gene="yjdG"
                /BC_number="3486..4511"
                /gene="yjdG"
                /note="function unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA22762.1"
                /db_xref="GI:2522011"
                /translation="MIIISYKPCNCGSDMAFDSTGSLSCSSCGRQNIISLPKENIAA
            RFSDDAKERQCKNGCAVLITTEATTATTCSEFCGAAITLADRLSGHLAARAKVLPFIS
            KOEAEQAFRRKCKKGLTPRGFMSADRIKISITGMVLPFMFIDLSVQVRAKCTOHAO
            YEEGDYICETEHEFAFRDINLDYLRIPVDASKRDELDKLEPSYSELKDFOTAY
            LAGYIAEKYNYDELEFPRAKERISSYIDSYIHSTSGTSTVNRDKHHTKNVSEY
            VLLPVMVSYDVERAHEIFAMNGQGTGVVKKPPIISGKVAAMFSGIAGTFLAKLVS
            LMGCGF"
            4497..4501
                /gene="yjdH"
                /BC_number="4497..4501"
                /gene="yjdH"
                /BC_number="4511..5275"
                /gene="yjdH"
                /BC_number="4511..5275"
                /note="function unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA22763.1"
                /db_xref="GI:2522012"
                /translation="MRGFFGKALFVVLAVFIMNPLLGIEAVRASLEQOHVYDRANLIS
            KAEIGLSEISALGAKRPTDFIITKSTNGEDIDYDGDFDRKSGKSTALITDM
            TNEVETIAEFKKEOYLDNSRLNSINNTISSDSENVEAEFTYIOLSVKDMGIRKG
            VANDNIEFPMNRKIDLAIIVGCIASVIMLYHAGGKVTYVNGSTYMDORTSDVIDQYOTY
            IHWYVIREKRPSSKDSGSGCTYKGTSTSGSRGSP"
            5296..5301
                /gene="yjdI"
                /BC_number="5296..5301"
                /gene="yjdI"
                /BC_number="5306..6277"
                /gene="yjdI"

```

```

/note="function unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA22764.1"
/db_xref="GI:2522013"
/translation="MSFRRNOLANVVEEPRDDMTFYMNREIKKSGRLIRPGOD
AVFLNNGRVEGVQDDGDYDIESEIIPLSTLKGKFGPNSGRARAVLVEVNRKEPVR
WGTQKAINIPAGMPCGMPIDRANGEPFVQDYISLIDKIAQKDYFVEDIKETIIS
ILDOLMKNTREGKDMFNLOANAFDIAKIGDLDLMOLISGMTVGTQISMSPYPO
EYDQMTKNASVGMGVNRYOISMTDGMASGMSGASDMDAGMNMNMNMNMNM
NOMNOMNOAOSSPOSTGSGSRNFCNPGCTGTGANGFCNPGCKLY"
6291..6323
/note="stem-loop"
/complement(6324..7365)
/gene="ydlJ"
/complement(6324..7349)
/gene="ydlJ"
/note="function unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA22765.1"
/db_xref="GI:2522014"
/translation="MNEGYIAGLLLTAGMIDELMTLMTESGAPITRCLSAWIMK
GCRKISGDHAKVLSMAGPILLCITLVITSLRSGVLIYSSDPSHMETOSKEPASM
SDRIYSGYVMEFTLNGDLAPNGLMKLVITLITACGLTTFTSYTLISVLSAVNOK
RSPQSVLSLGHGDEIVHNANMGKDPHDIDFLVASSSELGKLTQHNAPFLHFNK
STQHOESSIIAVLDEALTIFFKYGIPEQYQPNQHLIKARSISKNYLDTVHTATHP
AEQAPPEPDISKQSGIPALSKQTFQIVANSIKERQRLIGITQAGARKWPVEQAI
GNAYSPK"
complement(7358..7365)
/gene="ydlJ"
/922..9355
/gene="ydlK"
/7922..7927
/gene="ydlK"
/7354..9355
/gene="ydlK"
/note="putative"
/codon_start=1
/transl_table=11
/product="metabolite transport protein"
/protein_id="BAA22766.1"
/db_xref="GI:2522015"
/translation="NNKOGNOMSFRLTILVYFGGLFGYDVTGVLNGALPYMGEPDQ
LNTAFTEGLVTSLLFGALGAVFGKMSDFGRKNTLELAVFEISTIGTFEPN
VWMTISRFVLGAVGASVTYPAVLAEMSPVSRGRTONTLMTIVSOLAFAVNA
ILGTGDSHVMRMLVATSLPALFLFGMTMSPRMTVYSKGRKEDALRYAKIR
DEKRAAELOEIEFARKKEQLEKATFKDLSVFWRRIVYIGLAIVOQITGVNSIM
YGTETLRNSGFQTEALIGNIANGVISLAVTFWIMLGRGRRRLMTGLIGTTTA
LLIGLIFSLVBEGSPALPYVVLIVTVPFLAFOGALSPTWMLSEIFPLRGLGCMG
VYVFCLMVNFVAFSTFPIILAIGLSTFEFFIVGIGICSVLEVKRFLBETKGLSLEQ
LEENFRAYDSGAKKDSGAEVIG"
9360..9393
/note="stem-loop"
/complement(9403..10457)
/gene="ydlJ"
/complement(9403..10443)
/gene="ydlJ"
/note="putative"
/codon_start=1
/transl_table=11

```

Alignment scores:

Quality: 87.00 Length: 117
Ratio: 1.526 Gaps: 5
Percent Similarity: 48.718 Percent Identity: 26.496

Alignment block:

US-09-728-670-10 x AB007638 ..

Align seg 1/1 to: AB007638 from: 1 to: 16585

5 AsplysglytyrtyrlysglyaspaspalaSerTyrphecIuProth 21

```

||||| 11397 GATTAAGGC...CATGAGAGCGCTGCGATGTTCAATGCAATACGGAAG 11443
21 rglyProTyrLeuMetValasnValThrGlyValAspSerLygLyAsnG 38
11444 CGGTGATGACCTCTGCTGATTCATTAACAGATACGATGCGAT..... 11480
38 LuLeuLeuSerProHisTyrValGluPheProIleLeuProGlyThrThr 54
11481 ..CTGCGACACCAAGTATAT.....GAAATGCAAGGCG 11513
55 LeuThrLygGluLygIleGluTyrValGluThrPalaLeuAspAlaTh 71
11514 ATCATGAGGAGGAAGCTGACGACGATCCGCAATTCCTTGAT..... 11558
71 rAlaTyrLygGluPheArgValValGluLeuAspProSerAlaLygIleG 88
11559 .....CAAAACACG 11568
88 LuValThrTyrTyrAspLygAsnLygLygGluGluThrLygSerPhe 104
11569 AGTGCTCTATTATACGCGCATTAATGACAGATGAAAGAGCTTTT 11618
105 ProIleThrGluLygPheValAlaProAspLeuSerGluHisIleLy 121
11619 AAATTC.....GATCGAAGTCTATATAAA 11644
121 s 121
11645 A 11645
seq_name: gb_BA:BSUB0004

```

seq_documentation block:

LOCUS BSUB0004 213190 bp DNA linear BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 4 of 21): from 600701 to 813890.

ACCESSION 299107 AL009126
VERSION 299107.1 GI:2632866

KEYWORDS
SOURCE Bacillus subtilis.

ORGANISM Bacillus/Bacillales; Bacillus/Clostridium group;

REFERENCE 1 (bases 1 to 213190)

AUTHORS

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertolo, M.G., Bessieres, P., Bohlman, A., Borcher, S.,
Boriss, R., Boursier, L., Brans, A., Braun, M., Bricnell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J.,
Daniel, R.A., Denzot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
Foulger, D., Fritze, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghm, S.Y., Glaser, P., Goffeau, A., Golligly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer, Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kunita, K., Lapidus, A.,
Lainidis, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetle, D., Portolick, S., Prescott, A.M., Prescan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scottone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Seror, P., Shin, B.S., Sojdo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassatelli, A.,
Varti, A., Wamutic, R., Wedler, E., Wedler, H., Weltzenegger, T.,
Winters, P., Wipac, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,

Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H. and Danchin, A.
The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*
Nature 390 (6657), 249-256 (1997)

JOURNAL
MEDLINE
REFERENCE
AUTHORS
JOURNAL
TITLE
2 (bases 1 to 213190)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adachine@pasteur.fr Phone: +33 (0)1 45 88 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
source
Location/Qualifiers
1. .213190
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
33. .740
/gene="ydfS"
33. .740
/gene="ydfS"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12373.1"
/db_xref="GI:2632867"
/db_xref="SPTREMBL:P96697"
/translation="MIELEVYIRYVASGGLLIARILGKQTISOMTFIDPIAATITG
AIAAGLAYNTSIKPHNMAISFIVLTITFLISIKNRKLRKFFAGDPVYLIONGKI
LESNMRKRYTLDYLNQLRKEIFENIEVFALLETNGQJTVLRKQFRRHVTQDDM
IAVNOEQRPLEIIMDEGIENNNKONRLTESWLELRKDIRKKEIVYAVLLGNGD
IYVDYKDHISVPMDE"
complement(755. .1186)
/gene="ydfT"
complement(755. .1186)
/gene="ydfT"
complement(755. .1186)
/gene="ydfT"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12374.1"
/db_xref="GI:2632868"
/db_xref="SPTREMBL:P96698"
/translation="MDEKIRKWLITNEKYSDFWTNVLKKAPEHFFDEASTFYV
DEYODEEYNTIIVEMPGVEELTRILSKTQLIKGITFPFPAEMEVLRERYGEI
ERIIQPPAAETHLIQIOLNGLHISVPRQEVYAFMKGL"
complement(1199. .1441)
/gene="ydgA"
complement(1199. .1441)
/gene="ydgA"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12375.1"
/db_xref="GI:2632869"
/db_xref="SPTREMBL:P96699"
/translation="MPQIINTANIKINGVTONGNIDVGPVHNHNTANSKYEGANFSL
GDLSPSSLNNGTDSVSDODIGNSAPISNOI"
complement(1455. .1727)
/gene="ydgB"
complement(1455. .1727)
/gene="ydgB"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12376.1"
/db_xref="GI:2632870"
/db_xref="SPTREMBL:P96700"
/translation="MPSVIVINLYIKINSISGNSITIGEAAYNSPTNNQSGQTNS"

gene
CDS
EGDTSPTSEVMENFLDPVDNQTSGNSDTSNVAAPPIAPPIID"
2026. .2613
/gene="ydgC"
2026. .2613
/gene="ydgC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12377.1"
/db_xref="GI:2632871"
/db_xref="SPTREMBL:P96701"
/translation="MKSQGVESHKRLKKAANDESVRGTHDAKVSFIYKAKGFPDPSF
YLYFQSKREALFALITDFHSRVKRLTESLLENGLTFEDVSKRVLLAETVDFDLED
KDLTKIGFFLNPAPKQMKRDLAMVLEKNEAEORLQGFHSELDMETVACLVGMTEHL
TESFLTGIDKDPASLAEVVNLITGMLEKGDVR"
join(2610. .2954,3074. .3091)
/gene="ydgD"
2610. .2954
/gene="ydgD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12378.1"
/db_xref="GI:2632872"
/db_xref="SPTREMBL:P96702"
/translation="MISIMKVSIAFMLAGGIIRKRVYFQVEHMRHYQYPLMELTV
TGLIEIAGALATPAGIIMNRYAIGAGVLFVLMAGAIHMHFRARQSVYMAIQAMICL
IVSIIIMIGSYT"
3074. .3091
/gene="ydgD"
3117. .3590
/gene="ydgE"
3117. .3590
/gene="ydgE"
/gene="ydgE"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12379.1"
/db_xref="GI:2632873"
/db_xref="SPTREMBL:P96703"
/translation="MAFTLEDMPDEEFPFARQSVONPAKONITSGTVEKAPFEKSE
QAYENKIPNGRDSNHTFNNITNDGSRGMWLVADPLHPQEARVTSFGIYEARG
KGAQALQTLDRARELGAERLALHVFANETAVYLXOKGAYAMTNIMRRQLC"
complement(3705. .3722)
/gene="expZ"
complement(3705. .3722)
/gene="expZ"
complement(3705. .3722)
/gene="expZ"
complement(3750. .5393)
/gene="expZ"
complement(3750. .5393)
/gene="expZ"
/codon_start=1
/transl_table=11
/product="Atp-binding transport protein"
/protein_id="CAB12380.1"
/db_xref="GI:2632874"
/db_xref="SWISS-PROT:P39115"
/translation="MKEIVLTIVNVEVDQVFEKHVNASVOGDIIIGIIRKNGAGKS
TLHLHNDLAPAGOGILRDKILALVEOETAYSPADQPAEKULEKWHVPLRDFH
OISGGEELKARLAKGISADLILLDEPNHIDERSLOPLIOIKRYNNTVLIIVSDR
YFLDEATRKIWSLEDOTLIFKCNYSGYKFRKKRLTOORETEKQKAVEHIEAOMN
GLAWSERKAHQSTKKEGEYHVRVAKRTDAOIKSKORLEKLEKKAAEVTPPEYT
VRSIDTTHKTGRFLFVONVTKAFGERLFEKNANPTIQHGEVVAIIGNGSGKTTLL
NIIIGQETAGSVWSPSANGIYLTQEVDFPLEQTPPELEFENETFKARGHYONLMRH
LGFLAOWTEPIKHMWSGERVRIKMAVITLSEKDVLTILDEPTNHLDPREODELRS
OYSGTLLAVSHDRVPLEKTPNSKLVISNNGTIEKQLNDVPSERNRERELAKLETENOE
VLKLSFWTPNDGKYEKLDQAFNELIRKIKELDHQKQD"
complement(5428. .5450)
/gene="ydgF"
complement(5428. .7089)
/gene="ydgF"
complement(5678. .5703)
/gene="ydgF"

terminator
gene
terminator

CDS complement(5713..7089)
 /gene="ydgf"
 /function="unknown"
 /note="similar to amino acid ABC transporter (permease)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12381.1"
 /db_xref="GI:2632875"
 /db_xref="SWISS-PROT:P96704"
 /translation="MTDDMTKDNINOOTLRGKLNRIHIOIATGATGTLFGSGKS
 IHFAGPSILFRAHMTGICICLIRNSLSELLSNINHSFDFODYIGDMAATITGTT
 IWCWISIAADLTAVGLVQYWLPGVPGVPGILALILLINLATVRLGSELEWF
 ALKRVIALILVIGLVIRKGFSTSSGVSFTNLSHGLFPNGMRFTLSQMVVF
 AFVGIELVGLTAGETENPERKVIKPAINNIIRVLLFYGALLVIMSIYPMIDIINPSES
 PFVGVFAVGIVCAASIIINFEVLTSAASACNSAVFSTRMVSILAKDHPESMAKLT

alignment_scores:
 Quality: 87.00 Length: 117
 Ratio: 1.526 Gaps: 5
 Percent Similarity: 48.718 Percent Identity: 26.496

alignment_block:
 US-09-728-670-10 x BSUB0004 ..

Align seg 1/1 to: BSUB0004 from: 1 to: 213190

5 AsplysglylystyrylsyglyaspaspalaSertyrPhegluProth 21
 ||||| : : : : : ||||| : : : : :
 78922 GATTAAGGC...CATGGAAGGCTGCTGATGTTCTGCAATACCAAG 78966
 21 rgllyProtyrLeumetvalasnvalthrglyvalaspserlysglyasnG 38
 :
 78969 CGGTGATGAGCTGCTGTTCAATACAGATACGAT..... 79005
 38 luenuuserProhislyrvalgluPheproilelyserProglythrtrh 54
 ||||| : : : : : ||||| : : : : :
 79006 ..CTGCTGACAAAGTATAT.....GAAAATGACAGGTTG 79038
 55 leuthlysglylysiilegluTyryrvalgluTrpalaLeuaspalaTh 71
 :
 79039 ATCCATGAGAGAGAGAGCTGACGCTATCGGCATTGCTTGAT..... 79083
 71 ralatrylysglyluPheargvalajleuaspProseralalySileg 88
 :
 79084CAAAAACAGC 79093
 88 luvalthrtTyryrAsplyAsnlysylysglyluThrltyrSserPhe 104
 :
 79094 AGTGTCTTATATACGGCAATGAAACAGATGAATGAGGCTTTT 79143
 105 ProilethrglylysglyPhevalvalProaspLeuSergluHisilely 121
 :
 79144 AATGTG.....GATCTGACAGTCTCATATAAA 79169
 121 s 121
 79170 A 79170

seq_name: gb_pat:AX073718

seq_documentation_block:
 LOCUS AX073718 66 bp DNA linear PAT 06-FEB-2001
 DEFINITION Sequence 11 from Patent WO0104287.
 ACCESSION AX073718
 VERSION AX073718.1 GI:12710136
 KEYWORDS
 ORGANISM
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 66)
 AUTHORS Halkier,T., Pedersen,A.H. and Okkels,J.S.
 TITLE A method for preparing modified polypeptides
 JOURNAL Patent: WO 0104287-A 11 18-JAN-2001;

FEATURES Maxygen Aps (DK)
 location/Qualifiers
 source 1..66
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Primer"
 BASE COUNT 16 a 11 c 9 g 30 t
 ORIGIN

alignment_scores:
 Quality: 86.00 Length: 18
 Ratio: 5.059 Gaps: 0
 Percent Similarity: 94.444 Percent Identity: 94.444

alignment_block:
 US-09-728-670-10 x AX073718/rev ..

Align seg 1/1 to reverse of: AX073718 from: 1 to: 66

119 HsilelyAsnProglyPheasnleuileThrlsyvalIilegly 135
 ||||| : : : : : ||||| : : : : :
 65 CATATTAAAAACCTGATTCACCTTAAGACAAAGTTGTTATAGAAA 16

135 slys 136
 ||||
 15 GAAA 12

seq_name: gb_pat:AX073715

seq_documentation_block:
 LOCUS AX073715 66 bp DNA linear PAT 06-FEB-2001
 DEFINITION Sequence 8 from Patent WO0104287.
 ACCESSION AX073715
 VERSION AX073715.1 GI:12710133
 KEYWORDS
 ORGANISM
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 66)
 AUTHORS Halkier,T., Pedersen,A.H. and Okkels,J.S.
 TITLE A method for preparing modified polypeptides
 JOURNAL Patent: WO 0104287-A 8 18-JAN-2001;
 Maxygen Aps (DK)
 location/Qualifiers
 source 1..66
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Primer"
 BASE COUNT 18 a 9 c 9 g 30 t
 ORIGIN

alignment_scores:
 Quality: 85.00 Length: 18
 Ratio: 5.000 Gaps: 0
 Percent Similarity: 94.444 Percent Identity: 94.444

alignment_block:
 US-09-728-670-10 x AX073715/rev ..

Align seg 1/1 to reverse of: AX073715 from: 1 to: 66

119 HsilelyAsnProglyPheasnleuileThrlsyvalIilegly 135
 ||||| : : : : : ||||| : : : : :
 65 CATATTAAAAACCTGATTCACCTTAAGACAAAGTTGTTATAGAAA 16

135 slys 136
 ||||
 15 GAAA 12

seq_name: gb_pat:A05969

```

seq_documentation_block:
LOCUS      A05969          11225 bp      DNA      linear      PAT 11-JUN-1993
DEFINITION Complete sequence of the PMH23 insert (the '11.2kb BamHI
            fragment').
ACCESSION  A05969
VERSION    A05969.1  GI:412654
KEYWORDS
SOURCE     synthetic construct.
ORGANISM   synthetic construct
            artificial sequence.
REFERENCE  1 (bases 1 to 11225)
AUTHORS
TITLE      FOMLPOX VIRUS PROMOTERS
JOURNAL    Patent: WO 8903879-A 1 05-MAY-1989;
            Location/Qualifiers
FEATURES
            1..11225
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /complement(4821..5138)
            /gene="ORF5"
            /complement(4821..5138)
            /gene="ORF5"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA00470.1"
            /db_xref="GI:412654"
            /translation="MIRRNKALGSVMSDFIKTINEYDSNKEIKSEIDIKNSIL
            KEIDERYROEIKELCMIVDQLNQYKIIDNYSRYITEIRIQLLALKEENKCKEELT
            KKK"
            8025..8375
            /gene="ORF8"
            8025..8375
            /gene="ORF8"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA00471.1"
            /db_xref="GI:412656"
            /translation="MEBGRPRSSAVLWMLIPCGSIIVLSVFVITLSTRPVPPDIK
            ILYCKEGWGVNKNCYFFSEKNNKSLAVERCKMDGHLTSSKEEKFILRYKPG
            NHWIGIEKVDENGT"
            complement(8844..9686)
            /gene="ORF10"
            /complement(8844..9686)
            /gene="ORF10"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA00472.1"
            /db_xref="GI:412657"
            /translation="MKFKVRNTIKKMTDIIKIGINEYEMSKLLDVEVYINRSG
            FVNITRLCEKDFRNDKQLSSRRLLNTLKDNNKLDPIINIRIKIKINGEYVSO
            LLDVYIPWISPYVATRVSLMRVRCVALNIETEKIDHSQELQNIISKIDEYDR
            SIKDISNRKEIETSYSLSTYLLTKAERVLEKDYSMODIDNEDIRTEMDIAIE
            AEIEENRRRLSIISGIRKQHAEDRINISIKMLSGDSFNEIIVKIRDYIETAKPAAV
            NNYE"
BASE COUNT 3669 a 1720 c 1642 g 4193 t 1 others
ORIGIN
alignment_scores:
            Quality: 84.50      Length: 123
            Ratio: 1.408      Gaps: 5
            Percent Similarity: 48.780      Percent Identity: 25.203
alignment_block:
US-09-728-670-10 x A05969/rev ..
Align seg 1/1 to reverse of: A05969 from: 1 to: 11225
43 H1STYrValGluPheProIleLysProG1YThrThrLeuThrLysGluLy 59
|||||
11113 CATATTGTTGAAACCTATACATATAGCTGCTAATAGACAGAAAGTAA 11064
59 sile...GluTyrTyrValGluTyrPalaLeuAspAlaThrAla..... 72

```

```

|||||
11063 AATAGTAAATTCCTTATAGAAATACGCGCTGACATTAATTCAGAAAGACG 11014
73 .....TyrLysGlu 75
11013 GCGCGAATGCTAAATACCTATACATTACGCTATGAAAGTAAACGATCCG 10964
76 PheArgVal.....ValGluLeuAspProSerAlaLysIleG1 88
10963 TTATAGATTAAATAATATAAAGATATTATAGACACGCGCC..... 10922
88 uValThrTyrTyrAspLysAsnLysLysLysGluGluThrLysSerPheP 105
10921 .....GATATTACCAACAAACGCTTTTAACATAATACATCC 10885
105 rollerThrGluLysGlyPheValValProAspLeuSerGluHisIleLys 121
10884 CCTTATACGAACTAGCTTTATATACGACGACCTATTAGATTACATCAT 10835
122 AsnProGlyPheAsnLeu.....IlePth 129
10834 TCTAGAGAGAGCTATATATAATATAAGAAAGATGGGTAGAAATATATT 10785
129 rLysValValIleGluLys 135
10784 ACGAATATATATAGAACGA 10766

```

OM of: US-09-728-670-10 to: N.Geneseq_032802.* out_format : pfs
 Date: Sep 1, 2002 5:25 AM

About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlh
 -O=/cgm2_1/USPTO.spool/US09728670/r/unat.30082002.084914.5341/app_query.fasta.1.196
 -DB=N.Geneseq_032802 -QPM=fastap -SUFFIX=ring -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPOP=6.000 -FGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsu62
 -TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR SCORE=pct
 -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09728670.ecgnl_1_0 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPX -WAIT -THREADS=1

Search information block:

Query: US-09-728-670-10
 Query length: 136
 Database: N.Geneseq_032802.*
 Database sequences: 1736436
 Database length: 858457221
 Search time (sec): 210.660000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:ABA03011 +		715.00	1613.40	1.6e-81	41
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:ABA05083 +		715.00	1613.40	1.6e-81	41
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44270 +		715.00	1613.33	1.6e-81	41
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44269 +		711.00	1604.23	5.2e-81	41
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44271 +		711.00	1597.42	1.3e-80	80
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44272 +		711.00	1591.91	2.5e-80	1
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAQ44273 +		703.00	1586.18	5.3e-80	40
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44274 +		695.00	1567.84	5.6e-79	41
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44275 +		695.00	1555.11	7.9e-79	5
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44276 +		695.00	1558.56	1.8e-78	10
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ44277 +		693.00	1561.52	1.3e-78	49
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44278 +		689.00	1554.19	3.2e-78	41
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44279 +		684.00	1542.82	1.4e-77	41
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:AAQ44280 +		672.00	1506.58	1.4e-75	98
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44281 +		643.00	1450.33	1.9e-72	38
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:AAQ44282 +		620.00	1398.35	1.5e-69	37
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:AAQ44283 +		417.00	931.16	1.6e-43	63
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:AAQ44284 +		277.00	615.38	6.3e-26	45
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:AAQ44285 +		271.00	601.94	3.5e-25	48
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44286 +		264.00	585.71	2.8e-24	45
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44287 +		93.00	217.43	0.0009	4
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44288 +		92.00	214.69	0.0013	6
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44289 +		87.00	203.78	0.0053	6
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44290 +		86.00	201.50	0.0071	6
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44291 +		85.00	199.23	0.0095	6
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44292 +		84.50	145.44	9.40	11
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44293 +		84.00	136.95	0.0127	4
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAQ44294 +		82.50	132.69	4.821	2
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44295 +		81.50	148.89	6.04	4
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAQ44296 +		81.00	114.66	4.6e-8	10
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAQ44297 +		81.00	112.02	6.8336	13
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAQ44298 +		80.50	137.89	24.77	6
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:AAQ44299 +		80.00	147.74	7.00	33
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ44300 +		80.00	147.47	7.25	33
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAQ44301 +		80.00	114.65	4.6796	83
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAQ44302 +		80.00	113.83	5.4209	90
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAQ44303 +		79.50	160.04	1.45	85
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ44304 +		79.00	190.22	0.0301	4
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:AAQ44305 +		79.00	166.20	0.6555	43
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAQ44306 +		79.00	111.42	7.3810	9

/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ46743 - 78.00 128.28 84.88
 /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV72417 + 75.50 125.93 114.78
 /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ5113 + 75.00 181.12 0.0968
 /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AB134443 - 74.50 123.98 147.34
 /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAQ47096 + 74.00 141.24 16.10

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:ABA03011

seq_documentation block:

ID ABA03011 standard; DNA; 411 BP.
 AC ABA03011;
 XX
 DT 25-FEB-2002 (first entry)
 XX
 DE Sak glucokinase gene.
 XX
 KW Glucokinase; Sak; lysogenic conversion phage alpha;
 KW Staphylococcus aureus; fibrin; thrombolytic; haemorrhagic; ds.
 XX
 OS Staphylococcus aureus lysogenic conversion phage alpha.
 XX
 FH key Location/Qualifiers
 FT CDS 1..414
 FT /*tag= a
 FT /product= "glucokinase"
 FT /note= "CDS lacks an initiation codon"

CN1307136-A.

08-AUG-2001.

03-FEB-2000; 2000CN-0112674.

03-FEB-2000; 2000CN-0112674.

(SHAN-) SHANGHAI PLANT PHYSIOLOGY INST.

Wang J;

WPI: 2002-026897/04.

P-PSDB: AAK47900.

Recombinant Staphylococcus aureus glucokinase as thrombolytic medicine

Claim 1; Page 1 (Claims); 12pp; Chinese.

The invention relates to a glucokinase gene and a high expression engineered strain comprising a sak gene coding sequence obtained with DNA of lysogenic conversion phage alpha of Staphylococcus aureus through PCR amplification. The research of the applicant shows that recombinant glucokinase has specific effect on fibrin, has the features of powerful thrombolytic capacity and less haemorrhagic side effect and is safe and effective.

Sequence 411 BP; 162 A; 59 C; 80 G; 110 T; 0 other;

alignment_scores:

Quality: 715.00 Length: 136
 Ratio: 5.257 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-728-670-10 x ABA03011 ..
 Align seg 1/1 to: ABA03011 from: 1 to: 411

1 SerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTyr 17
 |||||
 1 TCAGTTCATTTCAGCAAGAAATATATATAAGGCGATGCGGAGTTA 50

	seq_name: /SISL/gcdate/geneseq/geneseqn-emb1/NA2002.DAT:ABA05083
seq_documentation_block:	
ID ABA05083 standard; DNA: 411 BP.	
XX	
AC ABA05083;	
XX	
DY 22-FEB-2002 (first entry)	
XX	
DE Human dipeptide aminopeptidase 28 coding sequence.	
XX	
KW Human; dipeptide aminopeptidase 28; cancer; nosohaemia; cytostatic;	
KM anti-HIV; immunosuppressive; antinflammatory; HIV infection;	
XX Immunological disease; inflammation; gene therapy; ds.	
XX OS Homo sapiens.	
XX	
FH Key	Location/Qualifiers
FT CDS	1..411
FT FT	/*tag= a
FT FT	/product= "dipeptide aminopeptidase 28"
FT FT	/partial
XX	/note= "the sequence contains no start codon"
CN1307129-A.	
XX	
PD 08-AUG-2001.	
XX	
PE 03-FEB-2000; 2000CN-0112673.	
XX	
PR 03-FEB-2000; 2000CN-0112673.	
PA (JINP-) JINPENG BIO TECH CO LTD SHANGHAI.	
XX	
PI Wang J, Wang S, Tong B;	
XX	
DR WPI: 2002-034896/05.	
DR P-PSDB: AAM47321.	
XX	
PT New polypeptide for treating malignant tumors and HIV infection,	
LT comprises the human dipeptide aminopeptidase -	

PS Claim 8; Page 1-2(claims); 11pp; Chinese.

XX
CC The present invention provides the protein and coding sequences of human
CC dipeptidase aminopeptidase 28. The sequences can be used in the treatment
CC of cancer, nosohemia, HIV infection, immunological diseases and
CC inflammation. The present sequence is the coding sequence of the
CC invention.
XX
SQ Sequence 411 BP; 162 A; 59 C; 80 G; 110 T; 0 other;

Alignment_scores:
Quality: 715.00 Length: 136
Ratio: 5.257 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
US-09-728-670-10 x ABA05083 ..

Align seg 1/1 to: ABA05083 from: 1 to: 411

1 SerSerSerPheAspLysGlyLeuTyrLysLysGlyAspAlaSerTyr 17
|||||
1 TCAGTCATCTTCGACAAAGAAAATATAAAAAAGCGCATGCACGTTA 50
|||||
17 rPhcgluProthrGlyProTyrLeuMetValAsnValThrGlyValAsps 34
|||||
51 TTTTGACCACCAACGCGCCGTATTGTATGGTAAATGTGACTGGAGTTGATA 100
|||||
34 eLySGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
101 GTAAAGGAAATGAATTCGATCCCTCATTAATGAGACTTCCTAATTAA 150
|||||
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
151 CCTGGCAGCACACTTACAAGAAAAAATTCATCTATGTCGAAATGGGC 200
|||||
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
201 ATTAGATGCGACAGCATATTAAGAAGTTAGAGTAGTTAATTAATGATCCA 250
|||||
84 ePAlaLysIleGluValrThyTyrAspLysAsnLysLysLysGluLu 100
|||||
251 GGCGAAAGATCGAAGTCACCTATTATGATAGAAATTAAGAAAAAAGAA 300
|||||
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLys 117
|||||
301 AGCAAGCTCTTCCCTATTAACAAAAAGGTTTTGTGTGCCGAGATTATC 350
|||||
117 rGlutHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleg 134
|||||
351 AGAGCATATTAAAAAACCTGATCAACTTAATTACAAAAGGTTGTTATAG 400
|||||
134 LuLysLys 136
|||||
401 AAAAGAAA 408

seq_name: /SIDB1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44270

seq_documentation_block:
ID AAQ44270 standard; DNA; 414 BP.
XX
XX AAQ44270;
XX AC
XX DT 03-DEC-1993 (first entry)
XX DE Staphylokinase SAK-STAR.
XX KW Signal sequence, plasmidogen activator; thrombosis; staphylokinase;
XX SAK; ss.
XX
XX Staphylococcus aureus strain 23.

XX MO9313209-A.
 PN 08-JUL-1993.
 PD 28-DEC-1992; 92WO-EP02989.
 PF 30-DEC-1991; 91DE-4143279.
 PR 22-JUN-1992; 92DE-4220516.
 PR 01-DEC-1992; 92DE-4240801.
 XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
 PA Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
 PI WPI: 1993-227325/28.
 DR P-PSDB; AAR39150.
 DR
 XX Staphylo-kinase (SAK) sequences lacking sequences for signal
 PT peptide(s) - for prodn. of proteins used as plasminogen
 PT activators in thrombosis treatment, and monoclonal antibodies
 PT against SAK
 XX
 PS Claim 3; Fig 3; 99pp; German.
 CC DNA encoding SAK lacking the signal peptide, is expressed
 CC intracellularly. This avoids the problem of fast degradation of the
 CC polypeptides or destruction of the host when expressed into the
 CC medium or into the periplasm respectively. High expression is
 CC possible and the chemically induced overprodn. is easy to handle.
 CC Also, the prods. are homogeneous.
 CC SAK-polypeptide derivs. are plasminogen activators for the
 CC treatment of thrombosis.
 CC
 XX Sequence 414 BP; 163 A; 59 C; 81 G; 111 T; 0 other;

alignment_scores:
 Quality: 715.00 Length: 136
 Ratio: 5.257 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-728-670-10 x AAQ44270 ..

Align seg 1/1 to: AAQ44270 from: 1 to: 414

```

1 SerSerSerPheaspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
4 TCAGATTTCATTCGACAAAGGAAATATATAAAAGCGATGACGCGATTA 53
17 rPhegluPProThnGlyProTyrLeuMetValAsnValThnGlyValAlaPS 34
|||||
54 TTTTGACCAACAGCGCCCTATTGATGGTAATGATGATGATGATGATA 103
34 eTylsGlyAsnGluLeuLeuSerProHlsTyrValGluPheProIleLys 50
|||||
104 GTAAAGGAATGAAATGCTATCCCTCATTTATGCGAGTTCTTATTA 153
51 ProGlyThrThleuthrLysGluLysIleGluTyrTyrValGluTTPAl 67
|||||
154 CCGGCGACTACACTTACAAAAGAAAAAATGAACTATGATGCAATGGCG 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
204 ATTAGATCGCAGCAGCATATTAAGAGTTTGAAGTGAATTAAGATCCAA 253
84 eAlaLysIleGluValThnTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCGCAAGATCGAAGTCACCTTATATGATTAAGATAAAGAAAAAGAGAA 303
101 ThLysSerPheProIleThnGluLysGlyPheValValProAspLeu 117
|||||

```

```

304 ACGAGCTCTTCCCTATACAGAAAAAGTTTGTGTGCCAGATTATC 353
117 rGluHIsIleLysAsnProGlyPheAsnLeuIleThrLysValValIleg 134
|||||
354 AGACATATTTAAAAACCTGGATTCACCTTATTAACAAAGTTGTTATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn_emb1/NA1993.DAT:AAQ44269

seq_documentation_block:

ID AAQ44269 standard; DNA; 414 BP.

XX AAQ44269;

DT 03-DEC-1993 (first entry)

DE Staphylokinase SAK-CphIC.

XX Signal sequence, plasminogen activator; thrombosis; staphylokinase;

KW SAK; ss.

OS Staphylococcus aureus phage phiC.

PN MO9313209-A.

PD 08-JUL-1993.

PE 28-DEC-1992; 92WO-EP02989.

PR 30-DEC-1991; 91DE-4143279.

PR 22-JUN-1992; 92DE-4220516.

PR 01-DEC-1992; 92DE-4240801.

XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.

PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;

DR WPI: 1993-227325/28.

DR P-PSDB; AAR39149.

PT Staphylo-kinase (SAK) sequences lacking sequences for signal

PT peptide(s) - for prodn. of proteins used as plasminogen

PT activators in thrombosis treatment, and monoclonal antibodies

PT against SAK

XX

PS Claim 3; Fig 2; 99pp; German.

CC DNA encoding SAK lacking the signal peptide, is expressed

CC intracellularly. This avoids the problem of fast degradation of the

CC polypeptides or destruction of the host when expressed into the

CC medium or into the periplasm respectively. High expression is

CC possible and the chemically induced overprodn. is easy to handle.

CC Also, the prods. are homogeneous.

CC SAK-polypeptide derivs. are plasminogen activators for the

CC treatment of thrombosis.

XX Sequence 414 BP; 161 A; 59 C; 83 G; 111 T; 0 other;

alignment_scores:

Quality: 711.00 Length: 136

Ratio: 5.228 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.265

alignment_block:

US-09-728-670-10 x AAQ44269 ..

Align seg 1/1 to: AAQ44269 from: 1 to: 414

```

1 SerSerSerPheaspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17

```

```

|||||
4 TCAAGTTCATTCCGACAAAGAAATATATAAAAGGCGCATGACGCGAGTTA 53
17 rphgluprothrglyprotyrleumetvalasnvalthrlyvalasps 34
54 TTTTGAACCAACAGCCCGCTATTGATGCTAAATGCTGAGAGTTGATG 103
34 erlysglyasnleuleuserprohistryvaliglupeproilelys 50
104 GTAAAGAAATGAATGCTATCCCTCATTTATGTCGAGTTCTCTATTA 153
51 ProGlyThrLeuThrLysGluLysIleGluTyrTyrValIGluTTPAL 67
154 CCGGAGCTACACTTACAAAGAAATAATGAAATGATGTCGATGGGC 203
67 aleuasplathralatryrlysglupeargvaliglupeasppros 84
204 ATTAGATCGCAGCAGCATTAAGAGTTAGAGTGTGAATTAATCCAA 253
84 eralalysilegluvalthrtyrtyrasplysasnllysglyglu 100
254 GCGCAAAAGATCGAAGTCACTTATATGATTAAGATAAGAAAGAA 303
101 Thrlysserpheproilethrlyglulysglyphevalproaspleuse 117
304 ACGAAGCTCTTCCCTATACAGAAAGCTTTGTTGTCGAGATTATTC 353
117 rgluhisilelyasnproglypheasnleuilethrlyvalalileg 134
354 AGAGCATATTAAAAACCTGATTCACACTTAATTACAAAGCTGTATTAG 403
134 lulysglys 136
404 AAAAGAAA 411

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:AAQ11813
seq_documentation_block:
ID AAQ11813 standard; DNA: 804 BP.
XX
AC AAQ11813:
XX
DT 05-AUG-1991 (first entry)
XX
DE Staphylococcus aureus promoter, Shine-Dalgarno and signal peptide
DE sequences.
XX
KM somatomedin; protein production; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT sig_peptide /*tag= a
FT promoter 193..222
FT RBS /*tag= b
FT /*tag= c
FT /standard_name= Shine-Dalgarno sequence
XX
PN JP03098595-A.
XX
PD 24-APR-1991.
XX
PF 11-SEP-1989; 89JP-0234874.
XX
PR 11-SEP-1989; 89JP-0234874.
XX
PA (TAIS ) TAISHO PHARMACEUT KK.
XX
DR WPI: 1991-167039/23.
XX
P-PSDB; AAR12137.
XX

```

```

PT Prepn. of peptide(s) - by construction of expression
PT vector, transformation of E.coli etc., culturing to secrete
PT peptide(s) and collecting peptide(s)
XX
PS Example: Fig 3; 15pp; Japanese.
XX
CC The sequence up to the EcoRI site at position 544 is ligated to a
CC synthetic Somatomedin C coding sequence to give a SMC-SMC construct.
CC Bacillus subtilis or E.coli can be transformed with such a
CC recombinant construct for expression and extracellular secretion of
CC somatomedin C. The SMC is secreted as a fusion protein with the
CC Staphylococcus gene product. See also AAQ11814.
XX
SQ Sequence 804 BP; 294 A; 89 C; 148 G; 273 T; 0 other;

alignment_scores:
Quality: 711.00 Length: 136
Ratio: 5.228 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.265

alignment_block:
US-09-728-670-10 x AAQ11813 ..
Align seg 1/1 to: AAQ11813 from: 1 to: 804

1 SerSerPheaspLyglyLyTyTyLyLyGlyAspAlaSerTy 17
|||||
394 TCAAGTTCATTCCGACAAAGAAATATATAAAAGGCGCATGACGCGAGTTA 443
17 rphgluprothrglyprotyrleumetvalasnvalthrlyvalasps 34
|||||
444 TTTTGAACCAACAGCCCGCTATTGATGCTAAATGCTGAGAGTTGATG 493
34 erlysglyasnleuleuserprohistryvaliglupeproilelys 50
|||||
494 GTAAAGAAATGAATGCTATCCCTCATTTATGTCGAGTTCTCTATTA 543
51 ProGlyThrLeuThrLysGluLysIleGluTyrTyrValIGluTTPAL 67
|||||
544 CCGGAGCTACACTTACAAAGAAATAATGAAATGATGTCGATGGGC 593
67 aleuasplathralatryrlysglupeargvaliglupeasppros 84
|||||
594 ATTAGATCGCAGCAGCATTAAGAGTTAGAGTGTGAATTAATCCAA 643
84 eralalysilegluvalthrtyrtyrasplysasnllysglyglu 100
|||||
644 GCGCAAAAGATCGAAGTCACTTATATGATTAAGATAAGAAAGAA 693
101 Thrlysserpheproilethrlyglulysglyphevalproaspleuse 117
|||||
694 ACGAAGCTCTTCCCTATACAGAAAGCTTTGTTGTCGAGATTATTC 743
117 rgluhisilelyasnproglypheasnleuilethrlyvalalileg 134
|||||
744 AGAGCATATTAAAAACCTGATTCACACTTAATTACAAAGCTGTATTAG 793
134 lulysglys 136
|||||
794 AAAAGAAA 801

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ85112
seq_documentation_block:
ID AAQ85112 standard; DNA: 1377 BP.
XX
AC AAQ85112:
XX
DT 08-MAY-2001 (first entry)
XX
DE DNA sequence of S. aureus staphylokinase.
XX

```

KM Immunogenicity: staphylokinase; variant; stability; ds.
 XX
 OS Staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 FT promoter 193..222
 FT RBS /*lag- a
 FT RBS 301..305
 FT CDS /*lag- b
 FT CDS 313..804
 FT sig_peptide /*lag- c
 FT sig_peptide 313..393
 FT mat_peptide /*lag- d
 FT mat_peptide 394..801
 FT /*lag- e
 FT /*note= "sequence AAB61908"
 XX
 PN W0200104287-A1.
 XX
 PD 18-JAN-2001.
 XX
 PF 06-JUL-2000; 2000MO-DK00371.
 XX
 PR 07-JUL-1999; 99DK-0000988.
 PR 27-AUG-1999; 99DK-0001196.
 PR 02-MAR-2000; 2000DK-0000339.
 PR 18-MAY-2000; 2000DK-0000804.
 XX
 PA (MAXY-) MAXYGEN APS.
 XX
 PI Halkier T, Pedersen AH, Okkels JS;
 XX
 DR WPI: 2001-138342/14.
 DR P-PSDB; AAB61908.
 XX
 PT Producing polypeptides with altered immunogenicity or improved
 PT stability, comprises expressing a diversified nucleotide sequence
 PT population and selecting polypeptides with altered immunogenicity or
 PT improved stability -
 XX
 PS Example 1; Page 74-75; 83pp; English.
 XX
 CC The invention relates to a method of altering immunogenicity and/or
 CC increasing stability of a polypeptide of interest. The method comprises
 CC (a) expressing a diversified population of nucleotide sequences encoding
 CC a polypeptide of interest; (b) screening the polypeptides expressed for
 CC function, immunogenicity and/or stability; and (c) selecting functional
 CC polypeptides with altered immunogenicity and/or increased stability. The
 CC method is used to improve the properties of polypeptides, in particular
 CC to alter the immunogenicity and/or increase the functional in vivo half-
 CC life of the polypeptide. The method uses a high throughput system that
 CC makes it possible to search several orders of magnitude more polypeptides
 CC than is possible by previously known approaches. This enhances the chance
 CC of finding the optimal variant from the many thousands of variants that
 CC may be produced. The present sequence represents a DNA encoding a
 CC S. aureus staphylokinase, used to exemplify the method of the invention.
 XX
 SO Sequence 1377 BP; 452 A; 184 C; 255 G; 486 T; 0 other;

alignment_scores:
 Quality: 711.00 Length: 136
 Ratio: 5.228 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.265

alignment_block:
 US-09-728-670-10 x AAC85112 ..
 Align seg 1/1 to: AAC85112 from: 1 to: 1377

1 SerSerSerPheAspLysGlyTyrLysGlyAspAlaSerTyr 17
 ||||||||||||||||||||||||||||||||||||||||
 394 TCAGTTCATTGCAAGAAATATATAAAGGCGCATGCGAGTTA 443

17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
 ||||||||||||||||||||||||||||||||||||||||
 444 TTTTGAAACCAACAGCGCCGATTTGATGATTAATGTCAGCTGATGATG 493
 34 eRlySGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
 :||||||||||||||||||||||||||||||||||||||
 494 GTAAAGGAAATGAATTCATCCCTCATATATGTCGAGTTCTCATTTAA 543
 51 ProGlyThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTyrP 67
 ||||||||||||||||||||||||||||||||||||||||
 544 CCTGGACTACACTTACAAAGAAATGAATGATCTATGTCGAAATGCGG 593
 67 AleuAspAlaThrAlaTyrLysGluPheArgValIleGluLeuAspPro 84
 ||||||||||||||||||||||||||||||||||||||||
 594 ATTGATGCCACGACCATATTAAGAGTTTAAGATTGAAATTGATCCAA 643
 84 eAlaLysIleGluValThrTyrTrpAspLysAsnLysLysGluGlu 100
 ||||||||||||||||||||||||||||||||||||||||
 644 GCCCAAGATCGAAGTCATTATATGATGAATAAGAAAGAAAGAA 693
 101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
 ||||||||||||||||||||||||||||||||||||||||
 694 ACGAAGCTTTCCCTATACGAAAGGTTGTGTGCGCAGATTATTC 743
 117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134
 ||||||||||||||||||||||||||||||||||||||||
 744 AGACCATATTTAAACCCTGATTCACCTTAATTACAAAGTTGTTATAG 793
 134 LuLysLys 136
 |||||||
 794 AAAAGAAA 801

seq_name: /stidsl/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAx27541

seq_documentation_block:
 ID AAX27541 standard; DNA; 408 BP.
 AC
 AC AAX27541;
 XX
 DT 27-MAY-1999 (first entry)
 XX
 DE S. aureus staphylokinase (Sak) encoding DNA.
 XX
 KW Staphylokinase; Sak; recombinant; myocardial infarction; cerebral;
 KW thrombolytic; fibrin; ss.
 KW hydrolytic; fibrin; ss.
 XX
 OS Staphylococcus aureus.
 XX
 PN W09904017-A1.
 XX
 PD 28-JAN-1999.
 XX
 PF 17-JUL-1998; 98MO-CN00129.
 XX
 PR 19-JUL-1997; 97CN-0105988.
 XX
 PA (BAIH/) BAI H.
 XX
 PI Ble L, Qu G, Wu Y, Xu G, Xu W, Zhang G, Zhang Q;
 XX
 DR WPI: 1999-132261/11.
 DR P-PSDB; AAY01232.
 XX
 PT Highly safe, novel recombinant staphylokinase (Sak) produced from
 PT high-expression engineered strain - as plasminogen activator, with
 PT very high hydrolytic activity to human fibrin, useful in treating
 PT thrombolytic diseases e.g. myocardial infarction
 XX
 PS Claim 2; Page 24; 51pp; Chinese.
 XX
 CC This DNA encodes a recombinant staphylokinase (Sak). The Sak-producing

CC Staphylococcus aureus S11.063 is deposited as GCMC No. 0353. The
 CC invention provides a method for constructing a Sak-producing engineered
 CC strain which comprises (a) screening Sak-producing Staphylococcus
 CC aureus; obtaining a required Sak gene by PCR (polymerase chain reaction)
 CC amplification with chromosomal DNA of Sak-producing S. aureus as
 CC template, with primers (AA27542-43) (b) introduction of the obtained
 CC DNA fragment into a plasmid selected from pUC19 and pBV220; and (c)
 CC transferring the recombinant plasmid into a host cell such as E. coli of
 CC DH5 alpha, TG1 or TG2 strains. The staphylokinase can be applied in
 CC treatment of myocardial infarction, thromboembolia diseases and arterial
 CC thrombosis including cerebral and pulmonary thrombi. Hydrolytic activity
 CC of the staphylokinase to human fibrin is very high.
 XX
 XX Sequence 408 BP; 156 A; 59 C; 83 G; 110 T; 0 other;
 SQ

alignment_scores:
 Quality: 703.00 Length: 136
 Ratio: 5.169 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.529

alignment_block:

US-09-728-670-10 x AA27541 ..

Align seg 1/1 to: AA27541 from: 1 to: 408

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
|||||
1 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGCGGAGCTTA 50
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyAlaSPS 34
|||||
51 TTTGAAACCAACAGAGCCGATTGATGATAAAGTGAAGTGAAGTGAAGT 100
34 eRlySGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|||||
101 GTAAAGAAATGAATTCCTATCCCTCGTATGTCGAGTTTCCTATTAA 150
51 ProGlyThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTPAI 67
|||||
151 CCGGAGACTACACTTACAAAGAAATTAATGATGATGCGAATGGGC 200
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
201 ATTAGATGCGACAGCATATAAAGATTAGAGTTGAATTTAGATCCAA 250
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
251 GCGCAAAAGATCGAAGTCATATTATGATAAAGATTAAGAAAGAAAGAA 300
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
301 ACGAAGCTTTCCTATACGAAAGAAAGGTTTGTGTCCAGATTATTC 350
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
351 AGACCATATTAATAAACCTGGATTCAACTTAATTAACAAAGGTTGATTG 400
134 LuLysLys 136
|||||
401 AAAAGAAA 408
```

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DMT:AAQ44268

seq_documentation_block:
 ID AAQ44268 standard; DNA: 414 BP.

XX AAQ44268;
 AC
 XX
 DT 03-DEC-1993 (first entry)
 XX
 DE Staphylokinase SAK42D.
 XX

KW Signal sequence, plasminogen activator; thrombosis; staphylokinase;
 KW SAK; ss.
 XX
 XX Staphylococcus aureus phage 42D.
 OS
 XX
 PN WO9313209-A.
 XX
 PD 08-JUL-1993.
 XX
 PF 28-DEC-1992; 92WO-EP02989.
 XX
 PR 30-DEC-1991; 91DE-414379.
 PR 22-JUN-1992; 92DE-4220516.
 PR 01-DEC-1992; 92DE-4240801.
 XX
 XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
 XX

PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
 XX WPI; 1993-227325/28.
 DR P-PSDB; MAR39148.
 XX

PT Staphylo-kinase (SAK) sequences lacking sequences for signal
 PT peptide(s) - for prodn. of proteins used as plasminogen
 PT activators in thrombosis treatment, and monoclonal antibodies
 PT against SAK
 XX
 XX Claim 2; Fig 1; 99pp; German.

CC DNA encoding SAK lacking the signal peptide, is expressed
 CC intracellularly. This avoids the problem of fast degradation of the
 CC polypeptides or destruction of the host when expressed into the
 CC medium or into the periplasm respectively. High expression is
 CC possible and the chemically induced overprod. is easy to handle.
 CC Also, the prods. are homogeneous.
 CC SAK-polypeptide derivs. are plasminogen activators for the
 CC treatment of thrombosis.
 CC
 XX
 XX Sequence 414 BP; 161 A; 59 C; 82 G; 111 T; 1 other;
 SQ

alignment_scores:
 Quality: 695.00 Length: 136
 Ratio: 5.148 Gaps: 0
 Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:

US-09-728-670-10 x AAQ44268 ..

Align seg 1/1 to: AAQ44268 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGCGGAGCTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyAlaSPS 34
|||||
51 TTTGAAACCAACAGAGCCGATTGATGATAAAGTGAAGTGAAGTGAAGT 103
34 eRlySGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|||||
104 GTAAAGAAATGAATTCCTATCCCTCGTATGTCGAGTTTCCTATTAA 153
51 ProGlyThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTPAI 67
|||||
154 CCGGAGACTACACTTACAAAGAAAGAAATTAATGATGATGCGAATGGGC 203
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
204 ATTAGATGCGACAGCATATAAAGATTAGAGTTGAATTTAGATCCAA 253
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
```


254 GCCCAAGATCGAGTCACTTATATGATAGATAGAAAAAGAGAA 303
 101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
 304 AGAAGTCTTCCCTATACAGAAAAAGTTTGTGTCCGATTTATC 353
 117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
 354 AGAGCATTTATAAAACCCGTGATCACTTATATACMAAGTTGTATAG 403
 134 LulysLys 136
 404 AAAAGAAA 411

seq_name: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH49196

seq_documentation_block:
 ID AAH49196 standard; DNA; 540 BP.

AC AAH49196;

DE 26-NOV-2001 (first entry)

XX E. coli Lac-YH1-Sak construct DNA.

KW L-form bacteria production; recombinant protein production;
 temperature stress; lactose permease helix 1; ds.

OS Escherichia coli.
 Synthetic.

XX Key Location/Qualifiers

FT misc_structure 1..129

FT /*tag= a

XX /*note= "Helix 1 region"

PM WO20016776-A2.

XX 13-SEP-2001.

PF 08-MAR-2001; 2001WO-EP02630.

PR 10-MAR-2000; 2000DE-1011617.

PR 12-MAR-2000; 2000DE-1011358.

XX (MOLE-) INST MOLEKULARE BIOTECHNOLOGIE EV.

PI Hoischen C, Gumpert J, Kujan JM, Fritsche C, Eliske G, Fahner B;

PI Steden S, Mueller HP;

DR WPI; 2001-582277/65.

DR P-PSDB; AAB86832.

XX Producing a modified L-form bacterial strain by culturing the strain in

PT complex medium and subjecting it to temperature stress, provides

PT bacteria useful to produce recombinant proteins

XX Claim 5; Fig 7a; 90pp; German.

XX This invention describes a novel method of producing a modified L-form

CC bacterial strain which comprises culturing a bacterial strain adapted to

CC a complex culture medium at alternate temperatures between 20 and 40

CC plusoc, and fermenting the strain when hydromechanical stress of the

CC cells increases. The modified L-form bacterial strains which are used to

CC produce recombinant proteins. This sequence encodes the Escherichia

CC coli derived construct LacYH1-Sak which is used to illustrate the method

CC of the invention.

CC Sequence 540 BP; 192 A; 87 C; 99 G; 162 T; 0 other;

XX alignment_scores: 695.00 Length: 136

Ratio: 5.148 Gaps: 0
 Percent Similarity: 99.265 Percent Identity: 97.794
 alignment_block:
 us-09-728-670-10 x AAH49196 ..

Align seg 1/1 to: AAH49196 from: 1 to: 540

1 SerSerSerPheAsnLysGlyLysValLysGlyAspAspAlaSerTy 17
 130 TCAGTTCATTCGACMAAGAAAATATMAAAAGCCATGACGAGTTA 179
 17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
 180 TTTGAACCAACAGAGCCGTTATGATGTFMAATGTGACTGAGATGATG 229
 34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
 230 GTAAAGAAATGAATTCCTATCCCTCGTTATGTCGAGTTCCATTAA 279
 51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrP 67
 280 CCTGGACTACACTTACAAAAGAAAATTAATCATCTATGTCGAAATGGC 329
 67 aLeuAspAlaThrAlaTyrTyrLysGluPheArgValValGluLeuAsp 84
 330 ATTAGATCGACAGCATTTAAAGATTAGAGTTGATTGAATTACATCAA 379
 84 eRAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
 380 GCGCAAGATCGAAGTCACTTATATGATAGTAAGTAAGTAAGTAAGTA 429
 101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
 430 AGAAGTCTTCCCTATACAGAAAAGGTTTGTGTCCGATTTATC 479
 117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
 480 AGAGCATTTATAAAACCCGTGATCACTTATATACMAAGTTGTATAG 529
 134 LulysLys 136
 530 AAAAGAAA 537

seq_name: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44276

seq_documentation_block:
 ID AAQ44276 standard; DNA; 1023 BP.

AC AAQ44276;

DE 03-DEC-1993 (first entry)

XX Staphylokinase SAK42D plasmid pDB17.

XX Signal sequence, plasminogen activator; thrombosis; staphylokinase;

XX SAK; ss.

OS Staphylococcus aureus phage 42D.

XX Key Location/Qualifiers

FT -35-signal

FT -10-signal

FT RBS

FT CDS

FT sig-peptide

FT mat-peptide

FT /*tag= a

FT /*tag= b

FT /*tag= c

FT /*tag= d

FT /*tag= e

FT /*tag= f

```

XX      /label= SAK42D
XX
XX      MO9313209-A.
XX
XX      08-JUL-1993.
XX
XX      28-DEC-1992;      92WO-BF02989.
XX
XX      30-DEC-1991;      91DE-A143379.
XX      22-JUN-1992;      92DE-4220516.
XX      01-DEC-1992;      92DE-4240801.
XX
XX      (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
XX
XX      Albrecht S, Behnke D, Guehrs K, Hartmann M, Schloft B;
XX      WPI: 1993-227325/28.
XX      P-PSDB: AAR39155.
XX
XX      Staphylo-kinase (SAK) sequences lacking sequences for signal
XX      peptide(s) - for produ. of proteins used as plasminogen
XX      activators in thrombolysis treatment, and monoclonal antibodies
XX      against SAK
XX
XX      Disclosure; Fig 9; 99pp; German.
XX
XX      DNA encoding SAK lacking the signal peptide, is expressed
XX      intracellularly. This avoids the problem of fast degradation of
XX      polypeptides or destruction of the host when expressed into the
XX      medium or into the periplasm respectively. High expression is
XX      possible and the chemically induced overprodu. is easy to handle
XX      Also, the produ. are homogeneous.
XX      SAK-polypeptide deriva. are plasminogen activators for the
XX      treatment of thrombolysis.
XX
XX      Sequence 1023 BP; 364 A; 123 C; 190 G; 346 T; 0 other;
XX
XX
XX
XX
XX      alignment_scores:
XX          Quality: 695.00      Length: 136
XX          Ratio: 5.148      Gaps: 0
XX      Percent Similarity: 99.265      Percent Identity: 97.794
XX
XX      alignment_block:
XX      US-09-728-670-10 x AAQ44276 ..
XX
XX      Align seg 1/1 to: AAQ44276 from: 1 to: 1023
XX
XX      1 SerSerSerPheaspLysGlyLysTyrLysLysGlyYAspAspAlaSerTyr 17
XX      TCAAATTCATTCGACAAAGAAAGAAATATATAAAAGGCATGACGCGAGTTA 474
XX
XX      17 rPheGluProthrGrgLProTyrLLeuMetValaAnValAThrgLValaAspS 34
XX      ITTTTTAACCAACAGGCCCGCTATTTGATGCGTAAGTAAGTACGTGGGTGAGG 524
XX
XX      34 eLysGlyYAsnGluLeuLeuSerProH1sTyrValaGluPheProIleTys 50
XX      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 574
XX      525 GTRAAAGAAATGAATGCTATCCCTCGTATATGCGAGTTTCCATTAA 574
XX
XX      51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValaGluTrrPal 67
XX      I|||||I|||||I|||||I|||||I|||||I|||||I|||||I|||||I||||| 624
XX      575 CCTGGGACTACACTTCACAAAAGAAAATTCGAACTACTATGTCGAAATGGGCG 624
XX
XX      67 aLeuAspAlaThrAlaTyrLysGluPheAcGValaValaGluLeuAspProS 84
XX      I|||||I|||||I|||||I|||||I|||||I|||||I|||||I|||||I||||| 674
XX      625 ATTACATCGCAGCAGCATTTAAAGATTAGAGTAGTTGAATTAAGATCCAA 674
XX
XX      84 eAlaLysIleGluValaThrTyrTyrAspLysAnLysLysGluGlu 100
XX      I|||||I|||||I|||||I|||||I|||||I|||||I|||||I|||||I||||| 724
XX      675 GCGCAAAAGATCGAAGTCACTTATTATGATTAAGATAAGAAAAAAGAA 724
XX
XX      101 ThrLysSerPheProIleThrGluLysGlyPheValaValaProaspLeu 117

```

```

725 ACGAAGCTCTTCCCATACGAAAAAGGTTTGTGTCCACAGATTATAC 774
|||||
117 TCUUHUHUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 134
|||||
775 AGAGCATATTAAAAACCCCTGGATTCAACTTAATTACAAAAGTTGTTATAG 824
|||||
134 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 136
|||||
825 AAAAGAAA 832

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AA094334

seq_documentation_block:
ID AA094334 standard; DNA: 492 BP.
XX
AC AA094334;
XX
DT 15-MAY-1996 (first entry)
XX
DE Recombinant staphylokinase gene SAK-1.
XX
KW Staphylokinase; Staphylococcus aureus; primer; PCR; amplification; lung;
KW prokaryote; expression vector; E.coli; chromatography; thrombus; brain;
KW peripheral blood vessel; heart attack; clotting; artery plug-in pipe;
KW operation; eye; elution; ss.
XX
OS Staphylococcus aureus.
XX
FH Key location/Qualifiers
FT CDS 427..429
FT /*tag=a
FT /transl_except= seq: ATA, a.a.: Leu
XX
PN MW0527048-A1.
XX
PD 12-OCT-1995.
XX
PE 03-APR-1995; 95MO-CN00025.
XX
PR 04-APR-1994; 94CN-0112105.
XX
PA (UYSH-) UNIV SHANGHAI MEDICAL.
XX
PI Song H;
XX
DR WPI: 1995-358627/46.
DR P-PSDB; AAR/5207.
XX
PT Prep'n. of recombinant staphylokinase - used to cure thrombus caused
PT by heart attack, blood clotting, etc.
XX
PS Examples: Page 11; 19pp; Chinese.
XX
CC The nucleotide sequence of the staphylokinase gene SAK-1 from
CC Staphylococcus aureus. The gene is amplified by the primers AA094332-3
CC and inserted into a prokaryotic expression vector such as pLY-4 to
CC produce the plasmid pSTE-SAK-1. The plasmid is used to transform E.coli
CC cells from where the kinase can be purified by a two-step
CC chromatographic method. The kinase is used to treat thrombus in lungs,
CC brains, and peripheral blood vessels after a heart attack, to prevent
CC clotting in artery plug-in pipes during operations, blood accumulating in
CC eyes and elution, etc.
XX
SO Sequence 492 BP; 186 A; 68 C; 91 G; 147 T; 0 other;

alignment_scores:
Quality: 693.00 Length: 136
Ratio: 5.133 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.059

alignment_block:

```

US-09-728-670-10 x AAQ94334 ..

Align seg 1/1 to: AAQ94334 from: 1 to: 492

```

1 SerSerSerPheaspLysGlyLysTyrlsGlyAspAspAlaSerTy 17
  |||||||
82 TCAGTTCATTCGACAAAGAAAATATATAAAAGGAGGAGCGAGCTTA 131
17 rPheGluProThrGlyProTyrlsMetValAsnValThrGlyValAsp 34
  |||||||
132 TTTTGACCAACAGAGCCCTATTATGATGTAATGAGTGGAGTTGATG 181
34 eTlrsGlyAsnGluLeuLeuSerProHisTyrlsValGluPheProIle 50
  |||||||
182 GTAAAGAAATGAAATGCTATCCCGCTTATGTCGAGTTCTTATAA 231
51 ProGlyThrThrlsLeuThrlsGlyLysGluTyrlsTyrlsValGlu 67
  |||||||
232 CCTGGGACTACCTTACAAAGAAAATTAATTAATGATGTCGAAATGG 281
67 AleuAspAlaThrAlaTyrlsGlyLysGluPheArgValGluLeuAsp 84
  |||||||
282 ATTAGATCGGAGACGACATATAAGAGTTTGAAGTGAATAGATCCAA 331
84 eTlAlaLysIleGluValThrTyrlsTyrlsAspLysAsnLysLysG 100
  |||||||
332 GCCCAAGATCGAAGCTCACTTATGATGAATGAAGAAAAGAACAA 381
101 ThrLysSerPheProIleThrGluLysGlyPheValAlaProaspLe 117
  |||||||
382 AGCAAGCTCTTCCCTATTAACAGAAAAGTTTGTGTCCAGATATATC 431
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrlsValValIle 134
  |||||||
432 AGAGCATATTAAAAACCTGGAATCAACTTAATTACAAAGTTGTATAG 481
134 LuLysLys 136
  |||||||
482 AAAAGAAA 489
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44273
seq_documentation_block:
ID AAQ44273 standard; DNA: 414 BP.
XX
AC AAQ44273;
XX
XX 03-DEC-1993 (first entry)
XX
DE Staphylokinase SAKM26C.
XX
KW Signal sequence, plasminogen activator; thrombosist; staphylokinase;
  SAK, ss.
XX
OS Staphylococcus aureus phage 42D.
XX
PN W09313209-A.
XX
PD 08-JUL-1993.
XX
PF 28-DEC-1992; 92MO-EP02989.
XX
PR 30-DEC-1991; 91DE-4143279.
PR 22-JUN-1992; 92DE-4220516.
PR 01-DEC-1992; 92DE-4240801.
XX
XX (MEDA-) MEDAC GES KLINISCHE SPEZIALLABORPARATE.
XX
PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
XX
DR WPI, 1993-227325/28.
XX
DR P-PSDB; AAR39153.
XX

```

PT Staphylo-kinase (SAK) sequences lacking sequences for signal
PT peptide(s) - for Prodn. of proteins used as plasminogen
PT activators in thrombosis treatment, and monoclonal antibodies
PT against SAK

PS Claim 5; Fig 6; 99p; German.

CC DNA encoding SAK lacking the signal peptide, is expressed
CC intracellularly. This avoids the problem of fast degradation of the
CC polypeptides or destruction of the host when expressed into the
CC medium or into the periplasm respectively. High expression is
CC possible and the chemically induced overprod. is easy to handle.
CC Also, the prods. are homogeneous.
CC SAK-polypeptide derivs. are plasminogen activators for the
CC treatment of thrombosis.

XX
SQ Sequence 414 BP; 158 A; 60 C; 81 G; 106 T; 9 other;

alignment_scores:

Quality:	589.00	Length:	136
Ratio:	5.142	Gaps:	0
Percent Similarity:	98.529	Percent Identity:	97.059

alignment_block:

US-09-728-670-10 x AAQ44273 ..

Align seg 1/1 to: AAQ44273 from: 1 to: 414

```

1 SerSerSerPheaspLysGlyLysTyrlsGlyAspAspAlaSerTy 17
  |||||||
4 TCAGTTCATTCGACAAAGAAAATATATAAAAGGAGGAGCGAGCTTA 53
17 rPheGluProThrGlyProTyrlsMetValAsnValThrGlyValAsp 34
  |||||||
54 TTTTAAACCAACAGGCCCGGATTTTGTGTGTAATGAGTGAATGATG 103
34 eTlrsGlyAsnGluLeuLeuSerProHisTyrlsValGluPheProIle 50
  |||||||
104 GTAAAGAAATGAGACTSYTRTCCCKGKTATGTCGAGTTCCATTAATA 153
51 ProGlyThrThrlsLeuThrlsGlyLysIleGluTyrlsValGluTPAl 67
  |||||||
154 CCTGGGACTACCTTACAAAGAAAATTAATTAATGATGTCGAAATGG 203
67 AleuAspAlaThrAlaTyrlsGlyLysGluPheArgValGluLeuAsp 84
  |||||||
204 ATTAGATCGAGACGATATTAAGAGTTTGAAGTGAATTAATAGATCCA 253
84 eTlAlaLysIleGluValThrTyrlsTyrlsAspLysAsnLysLysG 100
  |||||||
254 CGGCAAGATCGAAGTCACTTATATGATTAAGAAATTAAGAAAAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValAlaProaspLe 117
  |||||||
304 AGCAAGCTCTTCCCTATTAACAGAAAAGTTTGTGTCCAGATATATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrlsValValIle 134
  |||||||
354 AGAGCATATTAAAAACCTGATTCACCTTAATTACCAAGTTGTATAG 403
134 LuLysLys 136
  |||||||
404 AAAAGAAA 411
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44274
seq_documentation_block:
ID AAQ44274 standard; DNA: 414 BP.
XX
AC AAQ44274;
XX
XX 03-DEC-1993 (first entry)
XX

```

XX Staphylokinase SAKM26L.
 DE Signal sequence, plasminogen activator; thrombosin; staphylokinase;
 KM SAK; ss.
 OS Staphylococcus aureus phage 42D.
 XX
 PN W09313209-A.
 XX
 PD 08-JUL-1993.
 XX
 PF 28-DEC-1992; 92MO-EP02989.
 XX
 PR 30-DEC-1991; 91DE-4143279.
 PR 22-JUN-1992; 92DE-4220516.
 PR 01-DEC-1992; 92DE-4240801.
 XX
 PA (MEDA-) MEDAC GES KLINISCHE SPEZIALPREPARATE.
 XX
 PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
 XX
 DR WPI: 1993-227325/28.
 DR P-PSDB; AAR39154.
 XX
 PT Staphylo-kinase (SAK) sequences lacking sequences for signal
 PT peptide(s) - for prodn. of proteins used as plasminogen
 PT activators in thrombosis treatment, and monoclonal antibodies
 PT against SAK
 XX
 PS Claim 5; Fig 7; 99pp; German.
 XX
 CC DNA encoding SAK lacking the signal peptide, is expressed
 CC intercellularly. This avoids the problem of fast degradation of the
 CC polypeptide or destruction of the host when expressed into the
 CC medium or into the periplasm respectively. High expression is
 CC possible and the chemically induced overprod. is easy to handle.
 CC Also, the prods. are homogeneous.
 CC SAK-polypeptide derivs. are plasminogen activators for the
 CC treatment of thrombosis.
 CC
 XX Sequence 414 BP; 159 A; 60 C; 81 G; 105 T; 9 other;

alignment_scores:
 Quality: 684.00 Length: 136
 Ratio: 5.143 Gaps: 0
 Percent Similarity: 97.794 Percent Identity: 96.324

alignment_block:
 US-09-728-670-10 x MAQ44274 ..

Align seg 1/1 to: MAQ44274 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTy 17
 |||||
 4 TCAGATTCATTCGACAAAGGAAATATATAAAAGCGATGACGCGAGTTA 53
 |||||
 17 rPheGluPProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
 |||||
 54 TTTTGAACCAACAGCCCGCTATTGACGGTAAATGATCGATCGAGTATAG 103
 |||||
 34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
 ::|||
 104 GTAAAGAAATGARGCTSYTRCCCKGKATGTCGAGTTCCCTATTA 153
 |||||
 51 ProGlyThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTTPAl 67
 |||||
 154 CCGGGACTTACACTTACAAAGAAATTAATGATCATGTCGATGCG 203
 |||||
 67 AleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
 |||||
 204 ATTGATTCGACACCATATTAAGAGTTAGAGTGTGAATTAGATCCAA 253

84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
 |||||
 254 GCGCAAGATCGAAGTCACTTATATGATAGATTAAGATAAGAAAGAGNA 303
 |||||
 101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
 |||||
 304 ACGAAGTCCTTCCTATTAACAGAAAAAGGTTTGTTCGCCAGATTATTC 353
 |||||
 117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134
 |||||
 354 AGAGCATATTAAAAACCCCTGATTCACCTTAATTACMAAGTTGTTATAG 403
 |||||
 134 LuLysLys 136
 |||||
 404 AAAAGAAA 411

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:AAAT0606

seq_documentation_block:
 ID AAAT0606 standard; DNA: 991 BP.
 XX
 AC AAAT0606;
 XX
 DT 18-FEB-1991 (first entry)
 XX
 DE Sequence of staphylokinase gene from Staphylococcus aureus phage of
 DE serological gp. F esp. the phase ph1 42D.
 XX
 KW Enzyme; thrombolytic agent; bacterium; plasminogen; plasmin; ss.
 XX
 OS Staphylococcus aureus phage gp. F.
 XX
 FH Key Location/Qualifiers
 FT CDS 314..805
 FT /*tag= a
 FT -35_signal 194..199
 FT /*tag= b
 FT -10_signal 218..223
 FT /*tag= c
 FT RBS 302..307
 FT /*tag= d
 FT
 XX DD245444-A.
 PN
 XX
 PD 06-MAY-1987.
 XX
 PF 31-JAN-1986; 86DD-0286669.
 XX
 PR 31-JAN-1986; 86DD-0286669.
 XX
 PA (DEAK) AKAD WISSENSCHAFT DDR.
 XX
 PI Behnke D, Gerlach D, Adler B;
 XX
 DR WPI: 1987-264521/38.
 DR P-PSDB; AAP70381.
 XX
 PT New staphylokinase prodn. from heterologous prodn. cells -
 PT transformed with recombinant plasmids. useful as thrombolytic
 PT agent
 XX
 PS Disclosure: Fig 1, Page 10; 16pp; German.
 XX
 CC AAAT0606 is the SQ of the pref. gene for a new method of prodn. of
 CC staphylokinase. Using the new method, staphylokinase can be prepd.
 CC uncontaminated by toxic or antigenic materials and where Gram
 CC positive prodn. strains are used, it is excreted into the culture
 CC medium, facilitating recovery.
 CC
 XX Sequence 991 BP; 356 A; 118 C; 183 G; 334 T; 0 other;

alignment_scores: Quality: 672.00 Length: 136
 Ratio: 5.053 Gaps: 0
 Percent Similarity: 97.794 Percent Identity: 94.853

alignment_block:

US-09-728-670-10 x AAN70606 ..

Align seg 1/1 to: AAN70606 from: 1 to: 991

```

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
395 TCAAGTCTCTTCACAAAGAAATATATAAAGGCGATGACGAGTTA 444
17 PheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
445 TTTTGACCAACAGCGCCCTATTGATGAACTAGTACGTGAGTTGATG 494
34 eRLySGLyAsnGluLeuLeuSerProHisTyrValGluPheProIle 50
::|||
495 GTAAAGAAGAAATGAATTCATCCCTCGTTATGTCGAGTTTCATTA 544
51 ProGlyThrThrLeuThrLysGlyLysIleGlyTyrTyrValGluTri 67
|||||
545 CCTGGACTACACTTACAAAGAAATATGAACTACTGAAATGGGC 594
67 AleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
595 ATTAGATGCGACAGATATAAAGATTAGACTAGTGAATAGATCCAA 644
84 eRLAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
645 GCGCAAGATCGAAGTCACCTATATGATAGATAGAGAAAAAGAGAA 694
101 ThrLysSerPheProIleThrGlyLysGlyPheValValProAspLeu 117
|||||
695 AGGAGTCTTCTCCCTATACAGAAAAAGTTTGTGTCAGATTTATC 744
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
|||||
745 AGAGCATATATAAACCCTGATTCACCTTAATTAACAAAGTTGTTATAG 794
134 LuLysLys 136
|||||
795 AAAAGAAA 802
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44271
seq_documentation_block:
ID AAQ44271 standard; DNA; 384 BP.
XX
XX AAQ44271;
AC
XX
XX 03-DEC-1993 (first entry)
DT
XX
DE Staphylokinase SAKdeltaN10.
XX
KW Signal sequence, plasminogen activator; thrombosis; staphylokinase;
KW SAK, ss.
XX
OS Staphylococcus aureus phage 42D.
XX
XX MO9313209-A.
PN
XX
XX 08-JUL-1993.
PD
XX
XX 28-DEC-1992; 92WO-EP02989.
PF
XX
XX 30-DEC-1991; 91DE-4143279.
PR
XX 22-JUN-1992; 92DE-4220516.
PR
XX 01-DEC-1992; 92DE-4240801.
XX
XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
PA

```

```

XX
XX Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlotz B:
PI
XX WPI: 1993-227325/28.
DR
XX P-PDB; AAR39151.
XX
XX Staphylo-kinase (SAK) sequences lacking sequences for signal
PT peptide(s) - for Prodn. of proteins used as plasminogen
PT activators in thrombosis treatment, and monoclonal antibodies
PT against SAK
XX
XX Claim 4; Fig 4; 99p; German.
PS
XX
XX DNA encoding SAK lacking the signal peptide, is expressed
CC intracellularly. This avoids the problem of fast degradation of the
CC polypeptides or destruction of the host when expressed into the
CC medium or into the periplasm respectively. High expression is
CC possible and the chemically induced overprod. is easy to handle.
CC Also, the prods. are homogeneous.
CC SAK-polypeptide derivs. are plasminogen activators for the
CC treatment of thrombosis.
XX
XX
SO Sequence 384 BP; 146 A; 55 C; 78 G; 104 T; 1 other;

alignment_scores: Quality: 643.00 Length: 126
Ratio: 5.144 Gaps: 0
Percent Similarity: 99.206 Percent Identity: 97.619

alignment_block:
US-09-728-670-10 x AAQ44271 ..
Align seg 1/1 to: AAQ44271 from: 1 to: 384

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27
|||||
4 AAAGCGATGACGCGAGTATTGTAACCAACAGCCCGTATTGATGATGT 53
27 LAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisT 44
|||||
54 AAATGTGACTGAGTGTGATGATGATGATGATGATGATGATGATGATG 103
44 YrValGluPheProIleLysProGlyThrThrLeuThrLysGlyLysIle 60
|||||
104 ATGTGAGTTTCCCTATTAAACCTGGACTACCTTACAAAGAAATTAAT 153
61 GluTyrTyrValGluTyrPalaLeuAspAlaThrAlaTyrLysGluPhe 77
|||||
154 GAATRACTATGTCTGAAATGGCAATTGATGCGACACATATAAAGATTAG 203
77 GValValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAsp 94
|||||
204 AGTAGTTAATTAGATCCAGCCGCAAGATGAAAGTCACTTATATATATA 253
94 YAsnLysLysLysGluGluThrLysSerPheProIleThrGlyLysGly 110
|||||
254 AGAATTAAGAAAAAGAAAGAAAGAAAGTCTTCCCTATTAACGAAAAAG 303
111 PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsn 127
|||||
304 TTTGTTGTCGCCAGATTATATCAGACATATTAATAAACCTGATTCAC 353
127 uIleThrLysValValIleGlyLysLys 136
|||||
354 AATTACMAAGTTGTTATAGAAAAAGAAA 381
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44272
seq_documentation_block:
ID AAQ44272 standard; DNA; 372 BP.
XX
XX AAQ44272;
AC

```

```

XX 03-DEC-1993 (first entry)
XX Staphylokinase SAKdeltaN14.
DE Signal sequence, plasminogen activator; thrombosin; staphylokinase;
KW SAK; ss.
XX Staphylococcus aureus phage 42D.
OS
XX
XX MO9313209-A.
PN
XX
XX 08-JUL-1993.
PD
XX
XX 28-DEC-1992; 92WO-EP02989.
PF
XX
XX 30-DEC-1991; 91DE-4143279.
PR
XX 22-JUN-1992; 92DE-4220516.
PR
XX 01-DEC-1992; 92DE-4240801.
XX
XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPREPARATE.
XX
XX Albrecht S, Behnke D, Guehrs K, Hartmann M, Schloft B;
XX WPI; 1993-227325/28.
XX P-PSDB; AAR39152.
XX
XX Staphylo-kinase (SAK) sequences lacking sequences for signal
XX peptide(s) - for prodn. of proteins used as plasminogen
XX activators in thrombosis treatment, and monoclonal antibodies
XX against SAK
XX
XX Claim 4; Fig 5; 99p; German.
XX
XX DNA encoding SAK lacking the signal peptide, is expressed
XX intracellularly. This avoids the problem of fast degradation of the
XX polypeptides or destruction of the host when expressed into the
XX medium or into the periplasm respectively. High expression is
XX possible and the chemically induced overprod. is easy to handle.
XX Also, the prods. are homogeneous.
XX SAK-polypeptide derivs. are plasminogen activators for the
XX treatment of thrombosis.
XX
XX Sequence 372 BP; 141 A; 53 C; 74 G; 103 T; 1 other;
SO

```

alignment_scores:

Quality:	620.00	Length:	122
Ratio:	5.124	Gaps:	0
Percent Similarity:	99.180	Percent Identity:	97.541

alignment_block:

```

US-09-728-670-10 x AAO44272 ..
Align seg 1/1 to: AAO44272 from: 1 to: 372

```

```

15 AAlaserlyrphgeliuprotlnghlyprotylleuemetvalnsvalthrgl 31
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4 GCGAGTATTATTGAACACAGGCCCTATTGATGTAATGACTGG 53
31 YVAASerlyrsglyvaangliuleuSerProH1stYrValGluNhep 48
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54 AGTIGATGGTAAAGAAATGATGCTATGCCCTCGTATGTCGAGTTTC 103
48 roilelyrproglYthrThleuThrlysgliuleyllegluYrYrVal 64
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104 CTATTTAAACCTGGCAGTACACTTACAAAGAAAAAATTGATCTATGTC 153
65 GlutrrpalaleuaspAlatlnAlatYrlysgliupheargyValGluLe 81
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
154 GAAAGGCAATTAGATGCGACAGCATATAAGAGTTTGAAGTAGTGAATP 203
81 uasProSerAlalysilegluValThrYrYrAspYrlyAsnYrlysl 98

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
204 AGATCCAGCCGCAAGATCGAAGTCACTTATGATGAATAAGAGAAA 253
98 YsGluGluThrlyrSerPheProIleThrlysglyPheValYalPro 114
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
254 AAGAGAAAGCAAGTCTTCCCTATACAGAAAAAGTTTGTGTCCCA 303
115 AspLeuSerGluH1sIleYrAsnProGlyPheAsnleuIleThrlyrVa 131
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
304 GATTTATCAGACATATATAAAACCCTGATTCACCTTAATTACAAAGGT 353
131 YValIleGluYrlyr 136
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
354 TGTATAGAAAAAGAAA 369

```

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA1986.DAT: AAN60717

seq_documentation_block:

```

ID AAN60717 standard; DNA; 634 BP.
XX
XX AAN60717;
XX
XX 23-OCT-1991 (first entry)
XX
XX Fragment of sak gene in novel compound plasmid.
XX
XX Colibacillus; sak promoter; Psak; SIGsak; ss.
XX
XX Staphylococcus aureus.
XX
XX Key Location/Qualifiers
XX promoter 191..220
XX /*tag= a
XX /*label= Peak
XX RBS 299..303
XX /*tag= b
XX CDS 311..634
XX /*tag= c
XX sig_peptide 311..391
XX /*tag= d
XX /*label= SIGsak
XX
XX JP61135590-A.
XX
XX 23-JUN-1986.
XX
XX 05-DEC-1984; 84JP-0257258.
XX
XX 05-DEC-1984; 84JP-0257258.
XX
XX (HONS ) YAKULT HONSHA KK.
XX
XX WPI; 1986-267216/41.
XX DR P-PSDB; AAP60796.
XX
XX Compound plasmid - obt'd. by joining three specified DNA pieces to
XX PT colon bacterium plasmid vector.
XX
XX Disclosure; Fig 5; 17p; Japanese.
XX
XX The sequence is part of a compound plasmid construct comprising a
XX CC colibacillus promoter fragment, a S.aureus sak gene promoter (Psak)
XX CC ribosome binding site and signal peptide (SIGsak) and a restriction
XX CC enzyme recognition site at which an heterogene is inserted.
XX
XX Sequence 634 BP; 221 A; 63 C; 121 G; 229 T; 0 other;
SO

```

alignment_scores:

Quality:	417.00	Length:	81
Ratio:	5.213	Gaps:	0
Percent Similarity:	98.765	Percent Identity:	97.531


```

PA (TAIS ) TAISHO PHARMACEUT KK.
XX
DR WPI; 1991-167039/23.
DR P-PSDB; AAR12138.
XX
PT Prep'n. of peptide(s) - by construction of expression
PT vector; transformation of E.coli etc.; culturing to secrete
PT peptide(s) and collecting peptide(s)
XX
PS Example; Fig 4; 15pp; Japanese.
XX
CC This sequence encodes a SAK-SMC fusion protein comprising the
CC Staphylokinase signal peptide and part of the SAK mature protein
CC fused (via a linker) to somatomedin C. Expression of the fusion
CC protein in recombinant hosts is under control of staphylococcal
CC regulatory sequences.
CC See also AAQ11813.
XX
SQ Sequence 480 BP; 125 A; 102 C; 107 G; 146 T; 0 other;

alignment_scores:
    Quality: 271.00      Length: 51
    Ratio: 5.314        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 98.039

alignment_block:
US-09-728-670-10 x AAQ11814 ..
Align seg 1/1 to: AAQ11814 from: 1 to: 480

1 SerSerSerPheaspLySGlySTyLySLySGlyAspAspAlaSerTy 17
|||||
82 TCAGTTCATTCGACAAAGGAAATATATAAAGGCGGATGACGCGAGTTA 131
17 rPheGluProThrGlyProTyrluMetValAsnValThrGlyValAspS 34
|||||
132 TTTTGAACCAACAGGCCCGATTTGATGCTAAATGTGACTGGAGTTGATG 181
34 eRLySGLyAsnGluLeuLeuSerProHisTyrlValGluPheProIleLyS 50
:::
182 GTAAAGCAATGAATGCTATCCCTCATTTATGTCAGATTCTCTATTAA 231

51 Pro 51
111
232 CCT 234

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:AAQ24797
seq_documentation_block:
ID AAQ24797 standard; DNA: 495 BP.
XX
AC AAQ24797;
XX
DT 17-NOV-1992 (first entry)
XX
DE SAK/tryptstatin fusion protein.
XX
KW Staphylokinase; alkaline phosphatase; signal peptide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FH CDS 60..485
FT /*tag= a
FT /*label= SAK-tryptstatin_fusion_protein
FT /*tag= b
FT /*label= SAK_coding_region
FT /*tag= c
FT /*label= Linker
FT misc_feature 303..485

```

```

FT /*tag= d
FT /*label= tryptstatin_structural_gene
XX
XX JP04112791-A.
XX
PD 14-APR-1992.
XX
PE 31-AUG-1990; 90JP-0232268.
XX
PR 31-AUG-1990; 90JP-0232268.
XX
PA (TAIS ) TAISHO PHARM CO LTD.
XX
DR WPI; 1992-173147/21.
XX
PT Tryptstatin structural gene with specified DNA base sequence -
PT used with specified DNA to form expression vector, for commercial
PT prodn. of tryptstatin
XX
PS Disclosure; Fig 5; 18pp; Japanese.
XX
CC The sequences in AAQ24797-8 show two fusion constructs of SAK and a
CC synthetic tryptstatin structural gene. This sequence may be linked
CC to a DNA sequence encoding the signal peptide of the staphylokinase
CC gene originated from Staphylococcus aureus or of the alkaline
CC phosphatase gene originated from E.coli. This expression vector can
CC then be transformed into E. coli or Bacillus subtilis and the SAK-
CC tryptstatin fusion protein will be secreted extracellularly or in
CC the periplasm. It can then be recovered on a commercial scale.
XX
SQ Sequence 495 BP; 152 A; 91 C; 106 G; 146 T; 0 other;

alignment_scores:
    Quality: 264.00      Length: 53
    Ratio: 5.077        Gaps: 1
    Percent Similarity: 98.113    Percent Identity: 96.226

alignment_block:
US-09-728-670-10 x AAQ24797 ..
Align seg 1/1 to: AAQ24797 from: 1 to: 495

1 SerSerSerPheaspLySGlySTyLySLySGlyAspAspAlaSerTy 17
|||||
141 TCAGTTCATTCGACAAAGGAAATATATAAAGGCGGATGACGCGAGTTA 190
17 rPheGluProThrGlyProTyrluMetValAsnValThrGlyValAspS 34
|||||
191 TTTTGAACCAACAGGCCCGATTTGATGCTAAATGTGACTGGAGTTGATG 240
34 eRLySGLyAsnGluLeuLeuSerProHisTyrlValGluPheProIleLyS 50
:::
241 GTAAAGCAATGAATGCTATCCCTCATTTATGTCAGATTCTCTATTAA 290

50 sProGly 52
111111
291 ACCTGGG 297

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAC85121
seq_documentation_block:
ID AAC85121 standard; DNA: 66 BP.
XX
AC AAC85121;
XX
DT 08-MAY-2001 (first entry)
XX
DE S. aureus staphylokinase mutagenesis primer 2d (wt).
XX
KW Immunogenicity; staphylokinase; variant; stability; mutagenesis;
KW PCR primer; ss.
XX

```



```

OS Staphylococcus aureus.
XX
XX WO200104287-A1.
XX
XX 18-JAN-2001.
XX
XX 06-JUL-2000; 2000WO-DK00371.
XX
XX 07-JUL-1999; 99DK-0000988.
XX 27-AUG-1999; 99DK-0001196.
XX 02-MAR-2000; 2000DK-0000339.
XX 18-MAY-2000; 2000DK-0000804.
XX
XX (MAXY-) MAXYGEN APS.
XX
XX Halkier T, Pedersen AH, Okkels JS;
XX WPI; 2001-138342/14.
XX
XX Producing polypeptides with altered immunogenicity or improved
XX stability, comprises expressing a diversified nucleotide sequence
XX population and selecting polypeptides with altered immunogenicity or
XX improved stability -
XX
XX Example 2; Page 62; 83pp; English.
XX
XX The invention relates to a method of altering immunogenicity and/or
XX increasing stability of a polypeptide of interest. The method comprises
XX (a) expressing a diversified population of nucleotide sequences encoding
XX a polypeptide of interest; (b) screening the polypeptides expressed for
XX function, immunogenicity and/or stability; and (c) selecting functional
XX polypeptides with altered immunogenicity and/or increased stability. The
XX method is used to improve the properties of polypeptides, in particular
XX to alter the immunogenicity and/or increase the functional in vivo half-
XX life of the polypeptide. The method uses a high throughput system that
XX makes it possible to search several orders of magnitude more polypeptides
XX than is possible by previously known approaches. This enhances the chance
XX of finding the optimal variant from the many thousands of variants that
XX may be produced. Sequences AAC85113-121 represent PCR primers used in a
XX focused mutagenesis reaction towards introduction of lysine residues in
XX S. aureus staphylokinase.
XX
XX Sequence 66 BP; 18 A; 10 C; 9 G; 29 T; 0 other;
XX
XX
XX Alignment_scores:
XX Quality: 93.00 Length: 18
XX Ratio: 5.167 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-728-670-10 x AAC85121/rev ..
XX
XX Align seg 1/1 to reverse of: AAC85121 from: 1 to: 66
XX
XX 119 HisIleLysAsnProGlyPheAsnLeuIleThrLysValIleGluLys 135
XX |||||||
XX 65 CATATTAACCAACCGTGAATTCACCTTAATTACAAAGGTGTATATGAAAA 16
XX
XX 135 slys 136
XX ||||
XX 15 GAAA 12
XX
XX seq_name: /SID51/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAC85122
XX
XX seq_documentation_block:
XX ID AAC85122 standard; DNA; 69 BP.
XX
XX AC AAC85122;
XX
XX DT 08-MAY-2001 (first entry)
XX
XX DE S. aureus staphylokinase mutagenesis primer 3.

```

```

XX
XX Immunogenicity; staphylokinase; variant; stability; mutagenesis;
XX PCR primer; ss.
XX
XX Staphylococcus aureus.
XX
XX WO200104287-A1.
XX
XX 18-JAN-2001.
XX
XX 06-JUL-2000; 2000WO-DK00371.
XX
XX 07-JUL-1999; 99DK-0000988.
XX 27-AUG-1999; 99DK-0001196.
XX 02-MAR-2000; 2000DK-0000339.
XX 18-MAY-2000; 2000DK-0000804.
XX
XX (MAXY-) MAXYGEN APS.
XX
XX Halkier T, Pedersen AH, Okkels JS;
XX WPI; 2001-138342/14.
XX
XX Producing polypeptides with altered immunogenicity or improved
XX stability, comprises expressing a diversified nucleotide sequence
XX population and selecting polypeptides with altered immunogenicity or
XX improved stability -
XX
XX Example 3; Page 64; 83pp; English.
XX
XX The invention relates to a method of altering immunogenicity and/or
XX increasing stability of a polypeptide of interest. The method comprises
XX (a) expressing a diversified population of nucleotide sequences encoding
XX a polypeptide of interest; (b) screening the polypeptides expressed for
XX function, immunogenicity and/or stability; and (c) selecting functional
XX polypeptides with altered immunogenicity and/or increased stability. The
XX method is used to improve the properties of polypeptides, in particular
XX to alter the immunogenicity and/or increase the functional in vivo half-
XX life of the polypeptide. The method uses a high throughput system that
XX makes it possible to search several orders of magnitude more polypeptides
XX than is possible by previously known approaches. This enhances the chance
XX of finding the optimal variant from the many thousands of variants that
XX may be produced. Sequences AAC85122-123 represent PCR primers used in a
XX localised mutagenesis reaction to remove amino acid residues containing
XX attachment groups from a S. aureus staphylokinase.
XX
XX Sequence 69 BP; 30 A; 13 C; 10 G; 16 T; 0 other;
XX
XX
XX Alignment_scores:
XX Quality: 92.00 Length: 27
XX Ratio: 4.000 Gaps: 2
XX Percent Similarity: 85.185 Percent Identity: 81.481
XX
XX alignment_block:
XX US-09-728-670-10 x AAC85122 ..
XX
XX Align seg 1/1 to: AAC85122 from: 1 to: 69
XX
XX 83 ProSerIaIaLysIleGluValThrTyrTrpAspLysAsnLysLysG1 99
XX |||||||
XX 1 CCAACCGCTAAGATCGAAGTCACCTTATATGAT.....AATAAGA 41
XX
XX 99 uGluThrLysSerPheProIleThrGluLys 109
XX |||||||
XX 42 AGAAACG...TCTTCCCTATATACAGAAAA 69
XX
XX seq_name: /SID51/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAC85119
XX
XX seq_documentation_block:
XX ID AAC85119 standard; DNA; 66 BP.
XX
XX AC AAC85119;

```

```

xx 08-MAY-2001 (first entry)
xx
xx
xx S. aureus staphylokinase mutagenesis primer 2c.
xx
xx Immunogenicity: staphylokinase; variant; stability; mutagenesis;
xx PCR primer: ss.
xx
xx Staphylococcus aureus.
xx
xx WO200104287-A1.
xx
xx 18-JAN-2001.
xx
xx
xx 06-JUL-2000; 2000WO-DK00371.
xx
xx 07-JUL-1999; 99DK-0000988.
xx 27-AUG-1999; 99DK-0001196.
xx 02-MAR-2000; 2000DK-0000339.
xx 18-MAY-2000; 2000DK-0000804.
xx
xx (MAXY-) MAXYGEN APS.
xx
xx Halkier T, Pedersen AH, Okkels JS;
xx
xx WPI; 2001-138342/14.
xx
xx
xx Producing polypeptides with altered immunogenicity or improved
xx stability, comprises expressing a diversified nucleotide sequence
xx PT population and selecting polypeptides with altered immunogenicity or
xx PT improved stability -
xx
xx Example 2; Page 62; 83pp; English.
xx
xx
xx The invention relates to a method of altering immunogenicity and/or
xx CC increasing stability of a polypeptide of interest. The method comprises
xx CC (a) expressing a diversified population of nucleotide sequences encoding
xx CC a polypeptide of interest; (b) screening the polypeptide sequences expressed for
xx CC function, immunogenicity and/or stability; and (c) selecting functional
xx CC polypeptides with altered immunogenicity and/or increased stability. The
xx CC method is used to improve the properties of polypeptides, in particular
xx CC to alter the immunogenicity and/or increase the functional in vivo half-
xx CC life of the polypeptide. The method uses a high throughput system that
xx CC makes it possible to search several orders of magnitude more polypeptides
xx CC than is possible by previously known approaches. This enhances the chances
xx CC of finding the optimal variant from the many thousands of variants that
xx CC may be produced. Sequences AAC85113-121 represent PCR primers used in a
xx CC focused mutagenesis reaction towards introduction of lysine residues in
xx CC S. aureus staphylokinase.
xx
xx SQ
xx
xx Sequence 66 BP; 18 A; 11 C; 8 G; 29 T; 0 other;
xx
xx
xx alignment_scores:
xx
xx Quality: 87.00 Length: 18
xx Ratio: 4.833 Gaps: 0
xx Percent Similarity: 100.000 Percent Identity: 94.444
xx
xx alignment_block:
xx
xx US-09-728-670-10 x AAC85119/rev ..
xx
xx
xx Align seg 1/1 to reverse of: AAC85119 from: 1 to: 66
xx
xx 119 HsislslsyaasnProGlyPheasnLeuIllethrlsyalValileguly 135
xx |||||||||||||||||||
xx 65 CATATTAAAAACCTGGATTCACACTTAATTACAAAGGTGTGTTATAGAAA 16
xx
xx 135 slyS 136
xx ||||
xx 15 GAAA 12

```

```

seq_documentation_block:
ID AAC85120 standard; DNA; 66 BP.
XX
AC AAC85120;
XX
DT 08-MAY-2001 (first entry)
XX
DE S. aureus staphylokinase mutagenesis primer 2d.
XX
KW Immunogenicity; staphylokinase; variant; stability; mutagenesis;
KW PCR primer; SS.
XX
OS Staphylococcus aureus.
XX
PN MO200104287-A1.
XX
PD 18-JAN-2001.
XX
PF 06-JUL-2000; 2000MO-DK00371.
XX
PR 07-JUL-1999; 99DK-0000988.
XX
PR 27-AUG-1999; 99DK-0001196.
XX
PR 02-MAR-2000; 2000DK-0000339.
XX
PR 18-MAY-2000; 2000DK-0000804.
XX
PA (MAXY-) MAXYGEN APS.
XX
PI Halkier T, Pedersen AH, Okels JS;
XX
DR WPI; 2001-138342/14.
XX
PT Producing polypeptides with altered immunogenicity or improved
PT stability, comprises expressing a diversified nucleotide sequence
PT population and selecting polypeptides with altered immunogenicity or
PT improved stability -
XX
PS Example 2; Page 62; 83pp; English.
XX
CC The invention relates to a method of altering immunogenicity and/or
CC increasing stability of a polypeptide of interest. The method comprises
CC (a) expressing a diversified population of nucleotide sequences encoding
CC a polypeptide of interest; (b) screening the polypeptides expressed for
CC function, immunogenicity and/or stability; and (c) selecting functional
CC polypeptides with altered immunogenicity and/or increased stability. The
CC method is used to improve the properties of polypeptides, in particular
CC to alter the immunogenicity and/or increase the functional in vivo half-
CC life of the polypeptide. The method uses a high throughput system that
CC makes it possible to search several orders of magnitude more polypeptides
CC than is possible by previously known approaches. This enhances the chance
CC of finding the optimal variant from the many thousands of variants that
CC may be produced. Sequences AAC85113-121 represent PCR primers used in a
CC focused mutagenesis reaction towards introduction of lysine residues in
CC S. aureus staphylokinase.
CC
XX
SQ Sequence 66 BP; 16 A; 11 C; 9 G; 30 T; 0 other;
XX
XX
alignment_scores:
Quality: 86.00 Length: 18
Ratio: 5.059 Gaps: 0
Percent Similarity: 94.444 Percent Identity: 94.444
XX
alignment_block:
US-09-728-670-10 x AAC85120/rev ..
XX
Align seg 1/1 to reverse of: AAC85120 from: 1 to: 66
XX
119 HIsrIeLySaNPrOgLyPheASnLeuIeThrLySvAlIeGulY 135
|||||
65 CARATTTAAAAACCTGATTCACCTTAAAGACAAAGGTTGTTATGAAAA 16
135 sLyS 136
|||||

```

15 GAAA 12

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAC85117

seq_documentation_block:

ID AAC85117 standard; DNA: 66 BP.

XX AAC85117;

XX 08-MAY-2001 (first entry)

XX S. aureus staphylokinase mutagenesis primer 2a.

XX Immunogenicity: staphylokinase; variant: stability; mutagenesis;

XX PCR primer: ss.

XX Staphylococcus aureus.

XX WO200104287-A1.

XX 18-JAN-2001.

XX 06-JUL-2000; 2000WO-DK00371.

XX 07-JUL-1999; 99DK-0000988.

XX 27-AUG-1999; 99DK-0001196.

XX 02-MAR-2000; 2000DK-0000339.

XX 18-MAY-2000; 2000DK-0000804.

XX (MAXY-) MAXYGEN APS.

XX Halkier T, Pedersen AH, Okkels JS;

XX WPI; 2001-138342/14.

XX Example 2; Page 62; 83pp; English.

XX The invention relates to a method of altering immunogenicity and/or

XX increasing stability of a polypeptide of interest. The method comprises

XX (a) expressing a diversified population of nucleotide sequences encoding

XX a polypeptide of interest; (b) screening the polypeptides expressed for

XX function, immunogenicity and/or stability; and (c) selecting functional

XX polypeptides with altered immunogenicity and/or increased stability. The

XX method is used to improve the properties of polypeptides, in particular

XX to alter the immunogenicity and/or increase the functional in vivo half-

XX life of the polypeptide. The method uses a high throughput system that

XX makes it possible to search several orders of magnitude more polypeptides

XX than is possible by previously known approaches. This enhances the chance

XX of finding the optimal variant from the many thousands of variants that

XX may be produced. Sequences AAC85113-121 represent PCR primers used in a

XX focused mutagenesis reaction towards introduction of lysine residues in

XX S. aureus staphylokinase.

SQ Sequence 66 BP; 18 A; 9 C; 9 G; 30 T; 0 other;

alignment_scores:

Quality: 85.00

Ratio: 5.000

Percent Similarity: 94.444

Length: 18

Gaps: 0

Percent Identity: 94.444

alignment_block:

US-09-728-670-10 x AAC85117/rev ..

Align seg 1/1 to reverse of: AAC85117 from: 1 to: 66

119 HsIIcLysAsnProGlyPheAsnLeuIleThrLysValValIleGlyIu 135

|||||

|||||

|||||

65 CATATTAACCACTTAAGCTTCACTTATTACAAAGTGTGTATAGAAA 16

135 sLys 136

15 GAAA 12

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: AAN91695

seq_documentation_block:

ID AAN91695 standard; DNA: 11225 BP.

XX AAN91695;

XX 21-JAN-1991 (first entry)

XX Sequence of fowlpox virus (FPV) DNA insert in pMH23.

XX FPV vector; vaccine; chicken; fowl; ss.

XX Fowlpox virus.

XX Key Location/Qualifiers

XX CDS complement (4821..5138)

XX /*tag= a

XX /label=ORF5

XX /note="AAP96138"

XX CDS 8024..8375

XX /*tag= b

XX /label=ORF8

XX /note="AAP96139"

XX CDS complement (8844..9686)

XX /*tag= c

XX /label=ORF10

XX /note="AAP96140"

XX W08903879-A.

XX 05-MAY-1989.

XX 21-OCT-1988; 88WO-GH00922.

XX 23-OCT-1987; 87GB-0024885.

XX (NATR) NATIONAL RES DEV CORP.

XX Blains M, Boursnell MEG, Campbell JJA, Tomley FM;

XX WPI; 1989-150776/20.

XX P-PDB; AAP96138, AAP96139, AAP96140.

XX Fowl pox virus promoters -

XX used with foreign genes for providing recombinant vectors which

XX can be used for vaccinating responsive animals, esp. chickens

XX Example; P 24-31; 65pp; English.

XX Fowlpox virus (FPV) promoter DNA, for promoting the transcription of a

XX foreign gene inserted in a FPV vector is claimed. The promoters can

XX be used with foreign genes for providing recombinant vectors which

XX can be used for vaccinating responsive animals, esp. chickens. The FPV

XX promoters were identified by testing the 5' non-coding region of FPV

XX genes and selecting for strong promoters. FP4b, ORF10, ORF5 and ORF8

XX promoters are claimed. N = a nucleotide found to differ when

XX sequencing from different M13 clones of FPV.

SQ Sequence 11225 BP; 3670 A; 1719 C; 1643 G; 4192 T; 1 other;

alignment_scores:

Quality: 84.50

Ratio: 1.408

Percent Similarity: 48.780

Length: 123

Gaps: 5

Percent Identity: 25.203

PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Claim 1; SEQ ID NO 11437; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB16177), expressed DNA
 CC sequences (AB16177-AB16178) and the encoded proteins
 CC (AB16177-AB16178).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 24971 BP; 5336 A; 5636 C; 4884 G; 9115 T; 0 other;

alignment_scores:
 Quality: 82.50 Length: 141
 Ratio: 1.019 Gaps: 8
 Percent Similarity: 57.447 Percent Identity: 25.532

alignment_block:
 US-09-728-670-10 x ABL1998/rev ..

Align seg 1/1 to reverse of: ABL1998 from: 1 to: 24971

8 LysTyrLysLysGlyAspAspAlaSerTyrPheGlu..... 19
 20396 AGACATGAAAGAGGCGAGCACTGCAAGTTCGAAATGTGCAAGCGT 20347
 20ProThnGlyProTyrLeuMetValAsnVal...ThnGlyV 32
 20346 GGTGGCCCCCGGTGAAGAAACGCTTCATGAAATTAATTAATATCTAGG 20297
 32 AlaSerTyrGly...AsnGluLeuLeuSerProHisTyrValGluPhe 47
 20296 CAAATACGCAAGAGAGCACTGAAATTAACAACGTAAGTAAATA... 20250
 48 ProIleLysProGlyTyrThrLeuThrLysGluLysIleGluTyrTyrVa 64
 20249AGAAATTCGCAAGAAATTAATAAGGATGACCTGATCATATAT 20206
 64 LgluTrrAlaLeuAspAlaThrAlaTyrLysGluPheArgValAlaGlu 81
 20205 ACAGCAACTGATCATATGCC.....GAAATACCACTGACAGAAC 20168
 81 euAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnLysLys 97
 20167 TAGCAAAAATATGAAAATCGATGTAGTGAAGCAAAAACCGAA 20118
 98 LysGluGluThrLysSer...PheProIleThnGluLysGlyPheValVa 113
 20117 AAGCTTAAGGCGAGCAAGCAAAATTAATTAATGATGAGCGAACTCT 20068
 113 LPro...AspLeuSerGluHisIleLysAsnProGlyPheAsnLeuIleT 129
 20067 TCAGATTGAGTTCACAGCAT.....GAGCCACCAAAAATTAATAA 20024
 129 hrlsValValIleGluLysLys 136
 20023 CAAAAAACCAGGAGAAAGAA 20001

seq_name: /SIDS1/gcdata/geneseq/geneseqn_emb1/NA2001A.DAT: AAD10239

seq_documentation_block:
 ID AAD10239 standard; DNA; 4122 BP.

XX AAD10239;
 XX

DT 24-SEP-2001 (first entry)

XX Mouse long whey acidic protein (WAP) gene promoter fragment.

XX Mouse; whey acidic protein; WAP; promoter; protein C; milk; thrombin;
 KW transgenic animal; blood protein; Factor IX; FIX; Factor VIII; pig; cow;
 KW fibrinogen; FIB; von Willebrand's factor; sheep; goat; mice; rat; rabbit;
 KW ds.

XX Mus sp.

XX US6262336-B1.

XX 17-JUL-2001.

XX 28-MAY-1999; 99US-0321831.

XX 10-SEP-1992; 92US-0943246.

XX 11-JAN-1991; 91US-0638995.

XX 18-FEB-1994; 94US-0198068.

XX 17-MAY-1995; 95US-0443184.

XX 13-FEB-1998; 98WC-US02638.

XX 02-NOV-1998; 98US-0184063.

XX (AMNA-) AMERICAN NAT RED CROSS.

XX (VIRG) VIRGINIA TECH INTELLECTUAL PROPERTIES.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Lubon H, Drohan WN, Hennighausen L, Velandier WH;

XX WPI; 2001-450494/48.

XX DNA sequence that regulates the expression of heterologous gene

XX PT composed of a mouse whey acidic protein promoter extending upstream

XX PT from unique KpnI site in mouse whey acidic protein gene -

XX Example 11; Fig 9; 45pp; English.

XX The invention relates to an isolated DNA sequence which regulates the
 CC expression of heterologous gene, comprising a mouse long whey acidic
 CC protein (WAP) promoter fragment extending upstream from unique KpnI site
 CC in mouse whey acidic protein gene. The long WAP promoter is useful for
 CC the expression of a heterologous protein C in the mammary tissue of
 CC transgenic animals. Mouse WAP promoter fragment is highly efficacious in
 CC directing the production of high levels of active protein C in the milk
 CC of transgenic animals. The WAP promoter fragment is also useful in
 CC directing the production of high levels of polypeptides, particularly
 CC blood proteins such as factor IX (FIX), Factor VIII, human fibrinogen
 CC (hFIB), thrombin and von Willebrand's factor in the milk of a transgenic
 CC animal. The transgenic animals include mice, rats, rabbits, pigs, sheep,
 CC goats and cows. The present DNA sequence is mouse long whey acidic
 CC protein (WAP) gene promoter fragment.

XX Sequence 4122 BP; 1092 A; 1051 C; 994 G; 983 T; 2 other;

alignment_scores:
 Quality: 81.50 Length: 115
 Ratio: 1.509 Gaps: 6
 Percent Similarity: 46.957 Percent Identity: 29.565

alignment_block:
 US-09-728-670-10 x AAD10239/rev ..

Align seg 1/1 to reverse of: AAD10239 from: 1 to: 4122

42 ProHisTyrValGluPheProIleLysProGlyTyrThrLeuThrLysG 58
 3981 CCACACAGAGAGAGCGCCAGCTGCCAGCAAGCGCCACAAATGGA 3932
 58 ulysIleGluTyrTyrValGluTrrAlaLeuAspAlaThrAlaTyrLysG 75
 3931 GGCAAT.....TTTAAA 3918

```

75 lu.....PheArgValAlGluLeuAspProSerAlaLysIleGluVal 89
   :::::::::::::::::::::
3917 AGTCCTTTGTAAGCGGTGTCATTTTCCCACTATAAAGTAATCACTT 3868
   :::::::::::::::::::::
90 ThrTyrTyrAspLys..AsnLysLysLysGluGluThrLysSer..... 103
   :::::::::::::::::::::
3867 GTTATTATTAAGACCGGAAATGAAAGAGCAAAAGAAAGAAATAGG 3818
   :::::::::::::::::::::
104 .....PhePro 105
   :::::::::::::::::::::
3817 ATCAGTGAACAATGTGACAGTCACTTTGAGTGTGGGTGGAAACC 3768
   :::::::::::::::::::::
106 IleThrGluLysGly..PheValValProAspLeuSerGluHisIleLys 121
   :::::::::::::::::::::
3767 ATCAAGACAAAGCAATTCAGACCTCAAAATTTGTACACTCAATAGAA 3718
   :::::::::::::::::::::
122 Asn.....ProGlyPhe 125
   :::::::::::::::::::::
3717 AACAGACAGAGAGGTTTGAAGCAAGCTCGAGTCCGGGCTAC 3676
   :::::::::::::::::::::
seq_name: /SID1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAF22287

seq_documentation_block:
ID   AAF22287 standard; DNA; 103929 BP.
XX   AAF22287;
XX
XX   20-MAR-2001 (first entry)
XX
XX   BAC containing repeats from centromeres 1-4 #10.
XX
XX   Centromere; microsome; vector; ds.
XX
XX   Arabidopsis thaliana.
XX
XX   MO20055325-A2.
XX
XX   21-SEP-2000.
XX
XX   17-MAR-2000; 2000WO-US07392.
XX
XX   18-MAR-1999; 99US-0125219.
XX   01-APR-1999; 99US-0127409.
XX   18-MAY-1999; 99US-0134770.
XX   13-SEP-1999; 99US-0153584.
XX   17-SEP-1999; 99US-0154603.
XX
XX   (UYCH-) UNIV CHICAGO.
XX
XX   PI Preuss D, Copenhaver G, Keith K;
XX
XX   WPI; 2000-587529/55.
XX
XX   Recombinant DNA construct comprising a plant centromere, useful for
XX   producing stably inherited microsome which can serve as vectors for
XX   the construction of transgenic plant and animal cells.
XX
XX   Claim 102; Page 484-508; 1449pp; English.
XX
XX   The present invention relates to a recombinant DNA construct of a plant
XX   (Arabidopsis thaliana) centromere. The constructs are useful for
XX   producing stably inherited microsome which can serve as vectors for
XX   the construction of transgenic plant and animal cells expressing
XX   selected proteins such as hormones, enzymes, interleukins, clotting
XX   factors, cytokines, antibodies, and growth factors.
XX
XX   Sequence 103929 BP; 32987 A; 19310 C; 18547 G; 33005 T; 0 other;

```

alignment_scores:
 Quality: 81.00 Length: 138
 Ratio: 1.141 Gaps: 8

```

Percent Similarity: 51.449 Percent Identity: 28.986
alignment_block:
US-09-728-670-10 x AAF22287/rev ..
Align seg 1/1 to reverse of: AAF22287 from: 1 to: 103929
2 SerSerPheAspLys.....GlyLysTyrLysLysGlyAspAspAl 15
   :::::::::::::::::::::
16081 TCATCAATAGACAGCGGCGAGGTGGAATCTCAATAGAGGATAAAG 16032
   :::::::::::::::::::::
15 AsnTyrPheGluProThrGlyProTyrLeuMetValAsnValIThrGly 32
   :::::::::::::::::::::
16031 CATTTTAAAGAAACTGTGGGAAGTAC.....ANGAAAGAT 15994
   :::::::::::::::::::::
32 a1.AspSerLysGlyAsnGluLeuLeuSerProHisTyrValIupPhePr 48
   :::::::::::::::::::::
15993 TGGTCTGCTAAGCTCAATGACACACACTATGGGCTTACAGAACGCTTCA 15944
   :::::::::::::::::::::
48 o1LeLysProGlyThrThr..... 54
   :::::::::::::::::::::
15943 GACCCCTAATTGGCAGCAGCTCTTCAACCTCCTATAGGAAATCCTGTC 15894
   :::::::::::::::::::::
55 ..LeuThrLysGluLysIleGluTyrTyrValGluThrAlaLeu..... 68
   :::::::::::::::::::::
15893 ATTGGCTGTTGAA...CTCGAGTATAAAGCCATGCGGAGGTAAAGCTC 15847
   :::::::::::::::::::::
69 .....AspAlaThrAlaTyrLysGluPheArgValAlGluLeuAs 82
   :::::::::::::::::::::
15846 CTGAAGTTCGACATTAAACCGCCGAGGAGAAAGCGTTATATCAACTGAA 15797
   :::::::::::::::::::::
82 ProSerAlaLysIleGluValThrTyrTyrAspLysAsnLys..LysL 98
   :::::::::::::::::::::
15796 CGATCTCAACAGATTCGCTTAGAAGCCTATGAGATTCACAAATCTPAC 15747
   :::::::::::::::::::::
98 yseGluThrLysSerPhe.....ProIleThrGluLysGlyPhe 111
   :::::::::::::::::::::
15746 AGGAGCCAAACCAAGTCTTTCATGATAGAGATAGTCTCAAGAGATTTT 15697
   :::::::::::::::::::::
112 ValValProAsp 115
   :::::::::::::::::::::
15696 AAGGTTGCTGAT 15685
   :::::::::::::::::::::
seq_name: /SID1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAF22286

seq_documentation_block:
ID   AAF22286 standard; DNA; 134499 BP.
XX   AAF22286;
XX
XX   20-MAR-2001 (first entry)
XX
XX   BAC containing repeats from centromeres 1-4 #9.
XX
XX   Centromere; microsome; vector; ds.
XX
XX   Arabidopsis thaliana.
XX
XX   MO200055325-A2.
XX
XX   21-SEP-2000.
XX
XX   17-MAR-2000; 2000WO-US07392.
XX
XX   18-MAR-1999; 99US-0125219.
XX   01-APR-1999; 99US-0127409.
XX   18-MAY-1999; 99US-0134770.
XX   13-SEP-1999; 99US-0153584.
XX   17-SEP-1999; 99US-0154603.
XX
XX   (UYCH-) UNIV CHICAGO.
XX
XX   PI Preuss D, Copenhaver G, Keith K;

```



```

2391 ATCCGATGATCATTTGCGCTGAGCTCGAAGCAATCATCGTAACAACA 2440
71 .....ThralatryrlysgluPheargValVal 79
      ||||| ||||| |||||
2441 AGCAACTGATGCGGACACACACATCCCTATATGATTCATAGTAC 2490
80 GlutLeuAsp.....ProSerAlaLysIleGluValThrTy 91
      ||||| ||||| |||||
2491 GCCAAGGACAAAGGGGTGAMCCGCCAATTCGGAGTGCTACAGTTCAGAT 2540
91 rTyAspLysAsnLysLysGluGluThrLysSerPheProIleThrg 108
      ||||| ||||| |||||
2541 TTACACCAAAACAGACAGATGAGAG.....CCCAAGTTCT 2578
108 lutyrglyPheValValProAspLeuSerGluHisIleLysAsnProGly 124
      ||||| ||||| |||||
2579 CCCAGCAAGGTATATACCCCAATGTGATGAG.....AATGCCGGG 2619
125 PheAsn...LeuIleThrLysValValIleGluLysLys 136
      ||||| ||||| |||||
2620 CCMAACACTTGTGTACACTGTGTGCTCCGACAG 2658

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-embL/NA1991.DAT:AAQ14806

seq_documentation_block:
ID AAQ14806 standard; DNA; 3303 BP.
XX
AC AAQ14806;
XX
DT 10-FEB-1992 (first entry)
XX
DE B.churingiensis toxin/gp64 viral glycoprotein fusion gene pFAV10.
XX
KW chimeric; fusion protein; insecticide; AcNPV; Lepidoptera larvae;
KW midgut targeting; bacterial endotoxin; ss.
XX
OS Bacillus thuringiensis var. tenebriosis.
OS Autographa californica Nuclear Polyhedrosis Virus.
XX
PN WO9117254-A.
PD 14-NOV-1991.
XX
PF 02-MAY-1991; 91MO-US03008.
XX
PR 03-MAY-1990; 900S-0518575.
XX
PA (REGC ) UNITV OF CALIFORNIA.
XX
PI Sivasubramanian N, Federici A;
XX
DR WPI: 1991-353775/48.
DR P-PSDB: AARI5783.
XX
PT Extending host range or toxicity of insecticidal proteins - using
PT protein capable of binding to gut epithelium of insects
XX
PS Claim 24: Fig 16; 61pp; English.
XX
CC A polylinker was inserted into the XmnI restriction site at the
CC carboxyl terminus coding region of B.thuringiensis var. tenebriosis
CC (Bt) toxin. DNA encoding the gp64 viral membrane protein of AcNPV
CC was operably linked to the Bt toxin coding sequence via the
CC polylinker. The gp64 gene sequences act as midgut targeting
CC signals for bacterial endotoxins. Of three different Bt/gp64 gene
CC fusions that were constructed, pFAV10 was the longest. It was
CC transformed into E.coli BL21 and expressed. When tested against
CC heliothis virescens larvae, pFAV10 showed higher toxicity than
CC E.coli PSX12T expressing non-fusion Bt protein.
CC See also AAQ14807 and AAQ14808.
XX
SQ Sequence 3303 BP; 1077 A; 672 C; 724 G; 830 T; 0 other;

```

```

alignment_scores:
Quality: 80.00 Length: 147
Ratio: 1.250 Gaps: 6
Percent Similarity: 43.537 Percent Identity: 24.490

alignment_block:
us-09-728-670-10 x AAQ14806 ..

Align seg 1/1 to: AAQ14806 from: 1 to: 3303

4 PheAspLysGlyLysTyLysLysGlyAspAspAlaSerTy..... 17
      ||||| ||||| ||||| |||||
1663 TTCGATATAA...ACGATAAATAAGAGACACACATTATACGTAATTAATTCA 1709
18 .....PheGluProThrGlyProTyLysLeu 26
      ||||| ||||| |||||
1710 TTAATTAGCAAGTTTCACGACACACATTTCGAATTATACAGGAATTAACCTTAC 1759
26 etValAsnValThrGlyValAlaSerLysGlyAsnGluLeuLeuSerPro 42
      ||||| ||||| |||||
1760 AATAGGCGCTCACAGGATTA...AGTGCTGAGATTAAGTTTATATATAC 1806
43 HisTyValGluPheProIleLysPro..... 51
      ||||| ||||| |||||
1807 AAA...ATTGAATTGGCGATCGATCTCTAGAGTCGACCTGACGCCAAG 1853
52 .....GlyThrThrLeuThrLysGlyLysI 60
      ||||| ||||| |||||
1854 CTGCTCCGTACAAAGATTAAAACTTGACATTACCCGCCCAAGAAAGCCG 1903
60 lGluTyTyValGluThrPalalaLeuAspAlaThrAlaTyLysGluPhe 76
      ||||| ||||| |||||
1904 TGCMAAAGACGTGGAATATCACCATCTGTGAGACGAGCTACAAAGAAAC 1953
77 ArgValVal..... 79
      ||||| ||||| |||||
1954 GTGATTATCGGCTACAAAGGGGTACTACAGCGGTATGCGTACAAAGCGCG 2003
80 GluLeuAspProSerAlaLysIleGluValThrTyTyAspLysAsn 96
      ||||| ||||| |||||
2004 CTCGCTGATCCCAACACACACGCGTGAAGAAACATGAAGAACCGTCAATG 2053
96 yslLysLysGluGluThrLysSerPheProIleThrGluLys 109
      ||||| ||||| |||||
2054 TGGCAAAAGAGGATTTGCTCATGTGAGCATCAGCAGCAGCAG 2094

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-embL/NA1995.DAT:AAQ99223

seq_documentation_block:
ID AAQ99223 standard; cDNA to mRNA; 3394 BP.
XX
AC AAQ99223;
XX
DT 06-MAR-1996 (first entry)
XX
DE Japanese oyster transglutaminase cDNA.
XX
KW Japanese oyster; transglutaminase; gelling agent;
KW yoghurt; jelly; cheese; fish-paste; calcium ion activation; ds.
XX
OS Crassostrea gigas.
XX
FH Key Location/Qualifiers
FH CDS 305..2620
FH /*tag= a
XX
PN W09520662-A1.
XX
DR 03-AUG-1995.
XX
PF 30-JAN-1995; 95MO-JP00117.
XX

```


CC and is used in the method of the invention. The specification describes a method for modifying animal feed using galactanase. CC Galactanase enzymes degrade galactans. The galactanases of CC the invention are glycosyl hydrolases. The galactanase enzymes are CC useful for the modification of animal feed and in the textile, CC detergent, wine and juice and cellulose processing industries.

XX Sequence 891 BP; 261 A; 183 C; 263 G; 184 T; 0 other;

alignment_scores:

Quality: 79.50 Length: 116
Ratio: 1.242 Gaps: 7
Percent Similarity: 55.172 Percent Identity: 29.310

alignment_block:

US-09-728-670-10 x AAA59395 ..

Align seg 1/1 to: AAA59395 from: 1 to: 891

```

19 GIUProthrglyProtyrLeu...MetValAsnValThrglyValAspSe 34
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 GAACACACAGCGAGGATTTGCCAAGATGCGCAATTATTATGAAGAGAG 449
34 rlysglyasnGluLeuLeuSerProHisTyrValGluPheProIleLysP 51
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 CGGAGCGGTGAGTGTGCGCCCTCATTTTACCAAT...CCTGAACGG 496
51 rglYThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAla 67
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
497 CTGGAGGAGCTCAGACAAACAAAGTGATGATGATGCTG...TTTCT 543
68 LeuAspAlaThr.....AlaTyrLysGluPheArgVal 78
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 AGCCATGGACATTAACAAATTGACCTCCGTGGAAGCGCTGTAAGT 593
78 lValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspLysA 95
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
594 CATTG.....GTGCGGAGACATCTACACCTATACCGCTGAGA 631
95 sn.....LysLysLysGluGluThrLysSerPheProIleThrGluLys 109
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
632 ATACAGCACCAAAAGCGGACGCTTGCATATCCAAATTAATCT..... 675
110 GlyPheValValProAspLeuSerGluHisIleLysAsnProGlyPhe 125
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
676 .....GCAGTAAGAGATGTATATGAGCGACAGGCGAATACGGGCTTC 717

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAc85116

seq_documentation_block:

ID AAC85116 standard; DNA; 42 BP.
XX
AC AAC85116;
XX
DT 08-MAY-2001 (first entry)
XX
DE S. aureus staphylokinase mutagenesis primer 1d (vt).
XX
KW Immunogenicity; staphylokinase; variant; stability; mutagenesis;
KW PCR primer; ss.
XX
OS Staphylococcus aureus.
XX
PM WO200104287-A1.
XX
PD 18-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-DK00371.
XX
PR 07-JUL-1999; 99DK-0000988.
PR 27-AUG-1999; 99DK-0001196.
PR 02-MAR-2000; 2000DK-0000339.

PR 18-MAY-2000; 2000DK-0000804.

XX (MAXY-) MAXYGEN APS.

XX Halkier T, Pedersen AH, Okkels JS;

XX WPI; 2001-138342/14.

PT Producing polypeptides with altered immunogenicity or improved
PT stability, comprises expressing a diversified nucleotide sequence
PT population, and selecting polypeptides with altered immunogenicity or
PT improved stability -

XX Example 2; Page 62; 83pp; English.

CC The invention relates to a method of altering immunogenicity and/or
CC increasing stability of a polypeptide of interest. The method comprises
CC (a) expressing a diversified population of nucleotide sequences encoding
CC a polypeptide of interest; (b) screening the polypeptides expressed for
CC function, immunogenicity and/or stability; and (c) selecting functional
CC polypeptides with altered immunogenicity and/or increased stability. The
CC method is used to improve the properties of polypeptides, in particular
CC to alter the immunogenicity and/or increase the functional *in vivo* half-
CC life of the polypeptide. The method uses a high throughput system that
CC makes it possible to search several orders of magnitude more polypeptides
CC than is possible by previously known approaches. This enhances the chance
CC of finding the optimal variant from the many thousands of variants that
CC may be produced. Sequences AAC85113-121 represent PCR primers used in a
CC focused mutagenesis reaction towards introduction of lysine residues in
CC S. aureus staphylokinase.

XX Sequence 42 BP; 14 A; 9 C; 12 G; 7 T; 0 other;

alignment_scores:

Quality: 79.00 Length: 14
Ratio: 5.643 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-728-670-10 x AAC85116 ..

Align seg 1/1 to: AAC85116 from: 1 to: 42

```

10 LysLysGlyAspAspAlaSerTyrPheGluProThrGlyPro 23
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 AAAAAGCGCGATGACGCGAGTATTATTGAACCAACAGCCCGC 42

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:AAAN60718

seq_documentation_block:

ID AAN60718 standard; DNA; 437 BP.
XX
AC AAN60718;
XX
DT 23-OCT-1991 (first entry)
XX
DE Fragment of sak gene in novel compound plasmid.
XX
KW Colibacillus; sak promoter; Psak; SIGsak; ss.
XX
OS Staphylococcus aureus.
XX
PM Key Location/Qualifiers
PM promoter 191..220
FT /*tag- a
FT /label= Psak
FT RBS 299..303
FT /*tag- b
FT CDS 311..634
FT /*tag- c
FT sig_peptide 311..391
FT /*tag- d

seq_documentation_block:
ID AAS46743 standard; DNA; 14147 BP.
XX
XX AAS46743:
XX
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #467.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
XX cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX W0200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001MO-EP02955.
XX
XX 15-MAR-2000; 2000DE-1013847.
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AC.
XX
XX Olek A, Plepenbrock C, Berlin K;
XX
XX MPI; 2001-602752/68.
XX
XX
XX Fragments of chemically modified genes associated with tumour suppressor
XX genes and oncogenes, useful in designing primers and probes for
XX analysing diseases associated with cytosine methylation state e.g.
XX cancer
XX
XX
XX Claim 1; SEQ ID NO 467; 27pp: English.
XX
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and
XX oncogenes having a sequence taken from 536 (actually 533 since
XX numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX (SS) and sequences complementary to (SS). The nucleic acid may be a
XX peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX form part of a set of probes for detecting the cytosine methylation state
XX and/or single nucleotide polymorphisms and also to be used in an
XX array for analysing diseases associated with CpG dinucleotides e.g.
XX cancers and tumours. The probes can also be used in a method for
XX ascertaining genetic and/or epigenetic parameters for the diagnosis
XX and/or therapy of existing diseases or the predisposition to specific
XX diseases, by analysing cytosine methylations. The parameters may be
XX compared to another set of genetic and/or epigenetic parameters, the
XX differences serving as basis for diagnosis and/or prognosis events which
XX are disadvantageous to patients. The present sequence is one of the
XX 533 genomic sequences derived from tumour suppressor genes and
XX oncogenes.
XX
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 14147 BP; 3621 A; 225 C; 3116 G; 7185 T; 0 other;
XX

[illegible]

```

alignment_block:
US-09-728-670-10 x AAS46743/rev ..
Align seg 1/1 to reverse of: AAS46743 from: 1 to: 14147

```

```

28 AsnVal1ThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisTyr 44
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
13492 AATCCAAAAAATTAATTAATTCACCAACAC.....TTAAATAAACAAA 13449
44 rValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIleG 61
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
13448 AATAAATTTTCCCTAAACCT.....CCAAAAAAACGC 13414
61 LuTyrTyValGluTPrAlaLeuAspAlaThrAlaTyrLys.....74
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
13413 AACCTTACCAATFACCTTAATTTCACTAAACCAATTAATAAAGCTTAAAC 13364
75 .....GluPheArgValValGlu...LeuAspProSerAlaLysIleG1 88
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
13363 CTACAAAACCTATAAAATFACAAATATATTTTAAACAACTAAATTTAT 13314
88 uValThrTyTyAspLysAsnLysLysGluGluThrLysSerPhe. 104
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
13313 AATTAATTTATTAACAAACAAAAATAAAAACTAATTCACCTCTCATTTTA 13264
105 .....ProIleThrGluLysGly 110
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
13263 AATTAATTAATCTATCTCTACCATATGCAATATTTTCAAAAAAATT 13214
111 PheValValPro.....AspLeuSerGluHisIleLysAsnPr 123
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
13213 TTTTFTTTTACCAAAACCTTACAAAAACTACTACCTCCACGCTAAAGCTGC 13164
123 OGlyPhe 125
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
13163 TTCCTTC 13157

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NN1999.DAT:AAV72417
seq_documentation_block:
ID AAV72417 standard; DNA; 10220 BP.
AC
XX AAV72417:
XX
XX 03-AUG-1999 (first entry)
XX
XX P. luminescens pTetLux1 LuxCDABE DNA.
XX
XX Firefly, luciferase; tetracycline; transcriptional control; TetR; TetA;
XX tetracycline repressor; tetracycline promoter; Luminescence; luxCDABE;
XX Insect; Th10; medicine; dosage; cheese production; antibiotic; foodstuff,
XX allergy; lux C; lux D; lux A; lux B; lux E; ss.
XX
XX Photorhabdus luminescens.
XX
XX Key Location/Qualifiers
XX CDS 3634..5085
XX /*tag= a
XX /product= "lux C"
XX 5097..6020
XX /*tag= b
XX /product= "lux D"
XX 6069..7151
XX /*tag= c
XX /product= "lux A"
XX 7166..8149
XX /*tag= d
XX /product= "lux B"
XX 8256..9440
XX /*tag= e
XX /product= "lux E"
XX
XX W09925866-A1.

```


PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

•
•
•

OM of: US-09-728-670-10 to: Issued_Patents_NA:* out-format : pfs
 Date: Sep 1, 2002 4:30 AM

About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
 -O=/cgn2.1/USPTO.spool/US09728670/runat.30082002.084913.5318/app_query.fasta.1.196
 -DB=Issued_Patents_NA -QPMT=fastep -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
 -GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62
 -TRANS=human0.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pcit
 -THR_MAX=100 -THR_MIN=0 -ALIGN=50 -MODE=LOCAL -OUTPRM=pts
 -NORWEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09728670.ecgn1.1.45 -NCP=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPY -WAIT -THREADS=1

Search information block:

Query: US-09-728-670-10
 Query length: 136
 Database: Issued_Patents_NA:*
 Database sequences: 38353
 Database length: 122816752
 Search time (sec): 46.280000

score_list:

Sequence	Strd Orig	Zscore	Escore	len	Documentation	1..3e-87	414	1
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-5 +		715.00	1710.80	136	1.3e-87	414	1	9
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-852-299-5 +		715.00	1710.80	136	1.3e-87	414	1	9
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-3 +		711.00	1701.13	136	4.6e-87	414	1	9
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-852-299-3 +		711.00	1701.13	136	4.6e-87	414	1	9
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-1 +		695.00	1662.46	136	6.5e-85	414	1	9
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-852-299-1 +		695.00	1662.46	136	6.5e-85	414	1	9
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-16 +		695.00	1652.50	136	2.3e-84	1023	1	9
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-852-299-16 +		695.00	1652.50	136	2.3e-84	1023	1	9
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-11 +		689.00	1647.96	136	4.2e-84	414	1	9
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-852-299-11 +		689.00	1647.96	136	4.2e-84	414	1	9
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-13 +		687.00	1643.13	136	7.8e-84	414	1	9
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-852-299-13 +		687.00	1643.13	136	7.8e-84	414	1	9
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-7 +		643.00	1537.61	136	5.9e-78	384	1	9
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-852-299-7 +		643.00	1537.61	136	5.9e-78	384	1	9
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-9 +		620.00	1482.37	136	7.0e-75	372	1	9
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-852-299-9 +		620.00	1482.37	136	7.0e-75	372	1	9
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-09-321-831-1 -	84.50	150.61	1.06	11225	1.06	4122	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-321-831-1 -	81.50	154.39	0.6534	11225	0.6534	4122	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-09-251-645-11 +		73.50	110.61	179.31	0.7905	3394	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-251-645-11 +		73.00	148.91	1.32	1050	1050	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-480-481-3 +		73.00	148.91	1.32	1050	1050	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-480-481-3 +		73.00	148.91	1.32	1050	1050	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-525-654A-2 +		73.00	140.22	4.02	2313	2313	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-525-654A-2 +		73.00	140.21	4.03	8460	8460	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-469-005A-9 +		73.00	125.93	25.14	8519	8519	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-261-907-1 +		70.50	138.35	5.09	1576	1576	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-09-282-305-9 -		69.00	140.65	3.81	924	924	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-532-656-7 +		69.00	136.82	6.22	1308	1308	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-795-475-4 +		69.00	136.82	6.22	1308	1308	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-064-693A-22 +		69.00	136.49	6.49	1348	1348	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-09-064-693A-22 +		69.00	135.87	7.03	1278	1278	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-960-780-26 +		68.50	135.87	7.03	1278	1278	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-09-073-898-26 +		68.50	130.45	50.76	1485	1485	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-073-898-26 +		68.00	130.45	50.76	1485	1485	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-09-198-956-1 +		67.00	122.82	25.51	2290	2290	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-131-648-4 +		67.00	122.21	40.52	3178	3178	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-968-563-7 +		67.00	122.21	40.52	3178	3178	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-969-683A-7 +		67.00	122.21	40.52	3178	3178	1	Pat
/cgn2.6/ptodata/1/lna/5A.COMB.seq:US-09-297-928-3 +		66.50	133.29	9.78	1041	1041	1	Pat
/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-371-913A-10 +		66.50	133.29	9.78	1041	1041	1	Pat

/cgn2.6/ptodata/1/lna/6A.COMB.seq:US-09-313-930-1 + 66.00 124.33 30.86 2104
 /cgn2.6/ptodata/1/lna/5A.COMB.seq:US-09-147-208-26 - 66.00 107.58 264.73 9639
 /cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-978-741-5 - 66.00 105.84 330.71 11284
 /cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-852-481-1 + 65.50 126.39 23.70 1564
 /cgn2.6/ptodata/1/lna/6A.COMB.seq:US-08-232-079-1 + 65.50 120.48 30.58 2675

seq_name: /cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-256-261-5

seq_documentation block:

Sequence 5: Application US/08256261
 Patent No. 5801037
 GENERAL INFORMATION:
 APPLICANT: Behnke, Detlef
 APPLICANT: Schlotz, Bernhard
 APPLICANT: Albrecht, Sybille
 APPLICANT: G hrs, Karl-Heinz
 APPLICANT: Hartmann, Manfred
 TITLE OF INVENTION: Expression of signal-peptide-free
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Neave
 STREET: 1251 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10020

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/256,261

FILING DATE: 435
 CLASSIFICATION FOR SEQ ID NO: 5:
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 414 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "recombinant DNA"
 ORIGINAL SOURCE:
 ORGANISM: Staphylococcus aureus
 STRAIN: strain 23, genomic DNA

FEATURE:
 NAME/KEY: -
 LOCATION: 1..3
 OTHER INFORMATION: /note= "start codon"

FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..414

FEATURE:
 NAME/KEY: -
 LOCATION: 412..414
 OTHER INFORMATION: /note= "stop codon"

FEATURE:
 NAME/KEY: mat.peptide
 LOCATION: 4..411
 OTHER INFORMATION: /product= "mature protein"

US-08-256-261-5

alignment_scores:
 Quality: 715.00 Length: 136
 Ratio: 5.257 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-728-670-10 x US-08-256-261-5 ..

Align seg 1/1 to: US-08-256-261-5 from: 1 to: 414

```

1 SerSerSerPheaspLySGlyLysTyrLysLysGlyAspAspAlaSerTy 17
4 TCAGATTCATTCGACAAAGGAAATATATAAAAGGCGATGACCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
54 TTTTGAACCAACAGGCCGCTATTGATGATTAATGTGACTGGAGTTGATA 103
34 erLySGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
104 GTAAGGAATGATTCATTCCTCCCATTAATTCGAGTTTCCCTATTAA 153
51 ProGlyThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTPrAl 67
154 CCTGGGACTACCTACNAAAAGAAAATGATGACTATGTCGAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
204 ATTAGATGCGACAGCATATAAGATTAGATTAAGATTGAAATGATCCAA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGlyLysGlu 100
254 GCCCAAAAGATCGAAGTCATTAATGATTAAGATAAGAAAAAGAAAGAA 303
101 ThrLysSerPheProIleThrGlyLysGlyPheValValProAspLeu 117
304 ACGAAGCTTCCCTATTAACAGAAAAGGTTTGTGTGCCAGATTATTC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
354 AGAGCATATTAAACCCCTGGATTCACTTAATTACAAAGGTTGTTATAG 403
134 LuLysLys 136
404 AAAAGAAA 411

```

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-852-299-5

seq_documentation_block:

```

; Sequence 5, Application US/08852299
; Patent No. 6010897
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; TITLE OF INVENTION: staphylokinases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,299
; FILING DATE: 17-MAY-1997
; CLASSIFICATION: A35
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,261
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: strain 23, genomic DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
US-08-852-299-5

```

alignment_scores:

```

Quality: 715.00 Length: 136
Ratio: 5.257 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

US-09-728-670-10 x US-08-852-299-5 ..

Align seg 1/1 to: US-08-852-299-5 from: 1 to: 414

```

1 SerSerSerPheaspLySGlyLysTyrLysLysGlyAspAspAlaSerTy 17
4 TCAGATTCATTCGACAAAGGAAATATATAAAAGGCGATGACCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
54 TTTTGAACCAACAGGCCGCTATTGATGATTAATGTGACTGGAGTTGATA 103
34 erLySGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
104 GTAAGGAATGATTAATGATTCCTCCCATTAATGTCGAGTTTCCCTATTA 153
51 ProGlyThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTPrAl 67
154 CCTGGGACTACCTACNAAAAGAAAATGATGACTATGTCGAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
204 ATTAGATGCGACAGCATATAAGATTAGATTAAGATTGAAATGATCCAA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGlyLysGlu 100
254 GCCCAAAAGATCGAAGTCATTAATGATTAAGATAAGAAAAAGAAAGAA 303
101 ThrLysSerPheProIleThrGlyLysGlyPheValValProAspLeu 117
304 ACGAAGCTTCCCTATTAACAGAAAAGGTTTGTGTGCCAGATTATTC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
354 AGAGCATATTAAACCCCTGGATTCACTTAATTACAAAGGTTGTTATAG 403
134 LuLysLys 136
404 AAAAGAAA 411

```

```
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-256-261-3
seq_documentation_block:
; Sequence 3, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF INVENTIONS: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256,261
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phage phiC
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
US-08-256-261-3

alignment_scores:
Quality: 711.00 Length: 136
Ratio: 5.228 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.265

alignment_block:
US-09-728-670-10 x US-08-256-261-3 ..
Align seg 1/1 to: US-08-256-261-3 from: 1 to: 414
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTyr 17
|||||
4 TCAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGAGTTA 53
```

```
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTGTGAACCAACAGACCCGATATTGATGGTAATGTGACTGGAGTTGATG 103
|||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
104 GTAAGGAAATGAATTCGTATCCCTCATATATGCGAGTTCTTATTTAA 153
|||||
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPA 67
|||||
154 CCTGGGACTACACTTACAAAAGAAAAATGAACTATATGTCGAATGGCC 203
|||||
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspPro 84
|||||
204 ATTAGATGCGCAGACATATATAAGAGTTTAGAGTGTGAATTTAGATCAA 253
|||||
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCGCAAGATCGAAGTCATATATATATAGATTAAGAAAAAGAAAGAA 303
|||||
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
304 ACAGAGCTTTCCCTATATACAGAAAAAGTTTGTGTCCACAGTTTATC 353
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGACATATATTAACCCCTGATTCACACTTAATATACAAAGGTTGTATAG 403
|||||
134 luLysLys 136
|||||
404 AAAAGAAA 411

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-852-299-3
seq_documentation_block:
; Sequence 3, Application US/08852299
; Patent No. 6010897
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF INVENTIONS: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,299
; FILING DATE: 17-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,261
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
```

```

; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phage Phlc
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
US-08-852-299-3

```

```

alignment_scores:
    Quality: 711.00      Length: 136
    Ratio: 5.228         Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 99.265

```

```

alignment_block:
US-09-728-670-10 x US-08-852-299-3

```

```

Align seg 1/1 to: US-08-852-299-3 from: 1 to: 414

```

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTyr 17
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAAACCAACAGGCCCCGTATTGATGCTAAATGTGACTGGAGTTGATG 103
|||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
104 GTAAGCAAAATGAATGCTATCCCTCATATGTCGAGTTTCCATTTAA 153
|||||
51 ProGlyThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTyrPal 67
|||||
154 CCTGGACTACACTTACAAAAGAAAATTAAGTAATCTATGTCGAATGGGC 203
|||||
67 AleuAspAlaThrAlaTyrLysGluPheArgValAlaGluLeuAspPro 84
|||||
204 ATTGATGCCACAGCATATAAAGAGTTTACAGTAGTTGAATTAGATCCAA 253
|||||
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCCCAAGATCGAAGTCATTATGATGAATAAGAAAAGAAAGAA 303
|||||
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
304 ACGAAGCTTTCCTATTAACAGAAAAGGTTTGTGTCCCAATTTATC 353
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValAlaIle 134
|||||
354 AGACCATATTAAAAACCCGTGATTCACACTTAATTACAAAAGTTGTATAG 403
|||||
134 LuLysLys 136
|||||
404 AAAAGAAA 411

```

```

seq_name: /cgn2_6/ptodeta/1/lna/5A_COMB.seq:US-08-256-261-1

```

```

seq_documentation_block:
; Sequence 1, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:

```

```

; APPLICANT: Behnke, Detlef
; APPLICANT: Schlott, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; TITLE OF INVENTION: staphylokinases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (RPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256,261
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phage 42D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
US-08-256-261-1

```

```

alignment_scores:
    Quality: 695.00      Length: 136
    Ratio: 5.148         Gaps: 0
    Percent Similarity: 99.265    Percent Identity: 97.794

```

```

alignment_block:
US-09-728-670-10 x US-08-256-261-1

```

```

Align seg 1/1 to: US-08-256-261-1 from: 1 to: 414

```

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTyr 17
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAAACCAACAGGCCCCGTATTGATGCTAAATGTGACTGGAGTTGATG 103
|||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||

```

```

104 GTPAAGAAATGAATGCTATCCCTGTTATGTCGAGTTTCCTATTAA 153
51 ProGlyThrLeuThrLysGluLysIleGluTyrTyrValIleTrrpAl 67
|||||
154 CCTGGACACACCTTACAAAAGAAAAAATGAACTAGTGAATGGGC 203
67 aleuaspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
204 ATTAGATGCGACAGCATATAAGAGATTAGAGTAGTAATGATCAACA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysLysLysLys 100
|||||
254 GCGCAAGATCGAAGTCACCTTATTATGATTAAGATTAAGAAAAAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
304 AGAAGTCTTCCCTATACAGAAAAAGTTTGTGTGCCGATTTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAAAAAACCCTGATCAACTTAATACMAAGTTGTTATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411

```

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-852-299-1

seq_documentation_block:

```

; Sequence 1, Application US/08852299
; Patent No. 6010897
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,299
; FILING DATE: 17-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,261
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: phage 42D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414

```

```

; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat.peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
; US-08-852-299-1

```

```

alignment_scores:
  Quality: 695.00      Length: 136
  Ratio: 5.148        Gaps: 0
  Percent Similarity: 99.265  Percent Identity: 97.794

```

alignment_block:

US-09-728-670-10 x US-08-852-299-1 ..

Align seg 1/1 to: US-08-852-299-1 from: 1 to: 414

```

1 SerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTyr 17
|||||
4 TCAAGTTCATTCGACAAAGAAAATATATAAAGCGATGCGGATTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
54 TTTTGAAACCAACAGGCCCGTATTGTGATGCTAAATGCTACTGAGTTGATG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
104 GTPAAGAAATGAATGCTATCCCTGTTATGTCGAGTTTCTATTAA 153
51 ProGlyThrLeuThrLysGluLysIleGluTyrTyrValIleTrrpAl 67
|||||
154 CCTGGACACACCTTACAAAAGAAAAAATGAACTAGTGAATGGGC 203
67 aleuaspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
204 ATTAGATGCGACAGCATATAAGAGATTAGAGTAGTAATGATCAACA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysLysLysLys 100
|||||
254 GCGCAAGATCGAAGTCACCTTATTATGATTAAGATTAAGAAAAAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
304 AGAAGTCTTCCCTATACAGAAAAAGTTTGTGTGCCGATTTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAAAAAACCCTGATCAACTTAATACMAAGTTGTTATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-256-261-16

seq_documentation_block:

```

; Sequence 16, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free

```

```

? TITLE OF INVENTION: staphylokinases
? NUMBER OF SEQUENCES: 40
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Fish & Neave
? STREET: 1251 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: USA
? ZIP: 10020
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patentln Release #1.0, Version #1.30 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/256,261
? FILING DATE:
? CLASSIFICATION: 435
? INFORMATION FOR SEQ ID NO: 16:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1023 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: Staphylococcus aureus
? STRAIN: Phage 42D
? IMMEDIATE SOURCE:
? CLONE: Plasmid DB17
? FEATURE:
? NAME/KEY: -35_signal
? LOCATION: 224..229
? FEATURE:
? NAME/KEY: -10_signal
? LOCATION: 248..253
? FEATURE:
? NAME/KEY: -
? LOCATION: 332..336
? OTHER INFORMATION: /note="Shine-Dalgarno sequence"
? FEATURE:
? NAME/KEY: 519_peptide
? LOCATION: 344..415
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 416..832
? OTHER INFORMATION: /product="mature protein SAK42D"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 344..835
? US-08-256-261-16

alignment_scores:
? Quality: 695.00 Length: 136
? Ratio: 5.148 Gaps: 0
? Percent Similarity: 99.265 Percent Identity: 97.794

alignment_block:
US-09-728-670-10 x US-08-256-261-16 ..
Align seg 1/1 to: US-08-256-261-16 from: 1 to: 1023

1 SerSerSerPheaspLysGlyLysTyrLysGlyAspAspAlaSerTyr 17
|||||
425 TCAAGTTCATCGACAAAGAAATATATAAAGCGATGACGCGAGCTA 474
17 rPhgLuPrOthrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
475 TTTTGAACCAACAGCCCGTATTTGATGGTGAATGTAATGACTGAGATTGATG 524
34 eLysGlyAsnGluLeuSerProHsrTyrValGluPheProIleLys 50
::||| |||||

```

```

525 GTAAAGAAATGAATTCCTATCCCTGTTATGTCGACTTCCATTATAA 574
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
575 CCTGGACACTACACTACAAAGAAAAAATTGAATCTATGTCGATGGGC 624
67 AleuAspAlaThrAlaTyrLysGluPheArgValAlaGluLeuAspPro 84
|||||
625 ATTAGATCGACACGACATATAAAGATTAGACTGTAATTAATGATCCAA 674
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
675 GCGCAAGATCGAAGTCACTATTATGATAGAAATAGAAAAAGAGAA 724
101 rHlySerPheProIleThrGluLysGlyPheValProAspLeuSe 117
|||||
725 ACGAAGTCTTCCCTATACAGAAAAAGTTTGTTGCCAGATTATTC 774
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValAlaIleG 134
|||||
775 ACGACATATTAAAAACCCGTGATCACTTAATTAACAAGTTGTATAG 824
134 LuLysLys 136
825 AAAAGAA 832

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:us-08-852-299-16

seq_documentation_block:
? Sequence 16 Application US/08852299
? Patent No. 6010897
? GENERAL INFORMATION:
? APPLICANT: Behnke, Detlef
? APPLICANT: Schlotz, Bernhard
? APPLICANT: Albrecht, Sybille
? APPLICANT: G hrs, Karl-Helinz
? APPLICANT: Hartmann, Manfred
? TITLE OF INVENTION: Expression of signal-peptide-free
? NUMBER OF SEQUENCES: 40
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Fish & Neave
? STREET: 1251 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: USA
? ZIP: 10020
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patentln Release #1.0, Version #1.30 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/852,299
? FILING DATE: 17-MAY-1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/256,261
? FILING DATE:
? INFORMATION FOR SEQ ID NO: 16:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1023 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: Staphylococcus aureus
? STRAIN: Phage 42D
? IMMEDIATE SOURCE:
? CLONE: Plasmid DB17
? FEATURE:
? NAME/KEY: -35_signal

```



```
LOCATION: 224..229
FEATURE:
NAME/KEY: -10_signal
LOCATION: 248..253
FEATURE:
NAME/KEY: -
LOCATION: 332..336
OTHER INFORMATION: /note= "Shine-Dalgarno sequence"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 344..415
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 416..832
OTHER INFORMATION: /product= "mature protein SAK42D"
FEATURE:
NAME/KEY: CDS
LOCATION: 344..835
US-08-852-299-16
```

```
alignment_scores:
Quality: 695.00      Length: 136
Ratio: 5.148         Gaps: 0
Percent Similarity: 99.265      Percent Identity: 97.794
```

alignment_block:

US-09-728-670-10 x US-08-852-299-16 ..

Align seg 1/1 to: US-08-852-299-16 from: 1 to: 1023

```
1 SerSerSerPheaspLysGlyTyrLysGlyAspAspAlaSerTy 17
|||||
425 TCAGTTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 474
17 rphegluprothrglyprotyrleumetvalasnvalthrglyvalasps 34
|||||
475 TTTTGAACCAACAGCCCGTATTGATGTAATGTCAGTGGAGTGGAG 524
34 erlysglyasnGluLeuLeuSerProHisTyrValGluPheProIleTy 50
|||||
525 GTAAAGAAATGATGCTATCCCTCCCTGTTATGTCGAGTTTCTATTTAA 574
51 ProGlyThrThrLeuThrLysGlyLysIleGlyTyrTyrValGluTyrAl 67
|||||
575 CCTGGGACCTACCTACAAAAGAAAATGATGATGATGTCGAAAGGCG 624
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
625 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTGAATTAGATGCA 674
84 eraIalysIleGluValThrTyrTyrAspLysAsnLysLysGlyLysGlu 100
|||||
675 GCCCAAAAGATCGAAGCTATTATGATGATAGAAATAGAAAAAGAAAG 724
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
725 ACGAAGCTTTCCCTATTAACAGAAAAAGTTTGTGTGTCACAGATTAT 774
117 rgluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
775 AGAGCATATTAAAAACCTGATTCACCTTAATTAACAAAAGTTGTATAG 824
```

```
134 LuLysLys 136
|||||
825 AAAAGAAA 832
```

seq_name: /cgn2.6/ptodata/1/ina/5A.COMB.seq:US-08-256-261-11

seq_documentation_block:

; Sequence 11, Application us/08256261

; Patent No. 5801037

; GENERAL INFORMATION:

```
APPLICANT: Behnke, Detlef
APPLICANT: Schlot, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "recombinant DNA"
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: phage 42D
IMMEDIATE SOURCE:
CLONE: plasmid pMET5
FEATURE:
NAME/KEY: CDS
LOCATION: 1..414
FEATURE:
NAME/KEY: -
LOCATION: 1..3
OTHER INFORMATION: /note= "start codon"
FEATURE:
NAME/KEY: -
LOCATION: 412..414
OTHER INFORMATION: /note= "stop codon"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 4..411
OTHER INFORMATION: /product= "mature protein"
US-08-256-261-11
```

```
alignment_scores:
Quality: 689.00      Length: 136
Ratio: 5.142         Gaps: 0
Percent Similarity: 98.529      Percent Identity: 97.059
```

alignment_block:

US-09-728-670-10 x US-08-256-261-11 ..

Align seg 1/1 to: US-08-256-261-11 from: 1 to: 414

```
1 SerSerSerPheaspLysGlyTyrLysGlyAspAspAlaSerTy 17
|||||
4 TCAGTTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 53
17 rphegluprothrglyprotyrleumetvalasnvalthrglyvalasps 34
|||||
54 TTTTGAACCAACAGCCCGTATTGTTGTTAGTGAATGTCAGTGGAGTGGAG 103
```

```

34 ertlysglyasnlguleuserprohistryvalglupheproilleyls 50
:|||||
104 GTAAAGAAAGATGARTSYTRCCCKGKATGTGCGAGTTCCCTATTAAA 153
51 ProglyThrleuThrlysglyleuylleuylturytyrvalglutrpai 67
|||||
154 CTGGGACTACACTTACAAAAGAAAATTTGAATCTATGTCGATGGGC 203
67 aleuaspalaThrAlaTyrlYsglyupheargValaIgluleuaspProS 84
|||||
204 ATTAGATCGACACATATAAAGATTAGTAGTAGTTGAATTAGATCCAA 253
84 ertAlaIysIleGlyValThrTyTrYAspLysAnLysLysGlyGlu 100
|||||
254 GCGCAAAATCGAAGTCACCTATTATGATAAGATAAGAAAAGAA 303
101 ThrLysSerPheProIleThrGlyLysGlyPheValValProaspLeuSe 117
|||||
304 ACGAAGCTTTCCTATACAGAAAAGGTTTGTGTCCACATTTATC 353
117 rgluHsIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAAAAACCTGGATTCACACTTAATTACCAAGGTGTATAG 403
134 luLysLys 136
|||||
404 AAAAGAAA 411

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-852-299-11

```

seq_documentation_block:
; Sequence 11, Application US/08852299
; Patent No. 6010897
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schloft, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,299
; FILING DATE: 17-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,261
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phage 42D
; IMMEDIATE SOURCE:

```

```

; CLONE: Plasmid pMETS
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
US-08-852-299-11

```

alignment_scores:

```

Quality: 689.00 Length: 136
Ratio: 5.142 Gaps: 0
Percent Similarity: 98.529 Percent Identity: 97.059

```

alignment_block:

US-09-728-670-10 x US-08-852-299-11 ..

Align seg 1/1 to: US-08-852-299-11 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysTyrlYsglyAspAspAlaSerTy 17
|||||
4 TCAGTTCATTCGACAAAGAAATATATAAAGCGGATGCGCGATTA 53
17 rPheGluProThrGlyProTyrlLeuMeValAsnValThrGlyValAspS 34
|||||
54 TTTTGAACCAACAGCGCCGATTTGTGTAAATGTCGACTGAGATGAC 103
34 ertlysglyasnlguleuserprohistryvalglupheproilleyls 50
|||||
104 GTAAAGAAATGARTSYTRCCCKGKATGTGCGAGTTCCCTATTAAA 153
51 ProGlyThrleuThrlysglyleuylleuylturytyrvalglutrpai 67
|||||
154 CTGGGACTACACTTACAAAAGAAAATTTGAATCTATGTCGATGGGC 203
67 aleuaspalaThrAlaTyrlYsglyupheargValaIgluleuaspProS 84
|||||
204 ATTAGATCGACACATATAAAGATTAGTAGTAGTTGAATTAGATCCAA 253
84 ertAlaIysIleGlyValThrTyTrYAspLysAnLysLysGlyGlu 100
|||||
254 GCGCAAAAGATCGAAGTCACCTATTATGATAAGATAAGAAAAGAA 303
101 ThrLysSerPheProIleThrGlyLysGlyPheValValProaspLeuSe 117
|||||
304 ACGAAGCTTTCCTATACAGAAAAGGTTTGTGTCCACATTTATC 353
117 rgluHsIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAAAAACCTGGATTCACACTTAATTACCAAGGTGTATAG 403
134 luLysLys 136
|||||
404 AAAAGAAA 411

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-256-261-13

```

seq_documentation_block:
; Sequence 13, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schloft, Bernhard

```

```

APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
TITLE OF INVENTION: staphylokinases
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "recombinant DNA"
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: phase 42D
IMMEDIATE SOURCE:
CLONE: Plasmid pMETS
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 4..411
OTHER INFORMATION: /product= "mature protein"
FEATURE:
NAME/KEY: -
LOCATION: 1..3
OTHER INFORMATION: /note= "start codon"
FEATURE:
NAME/KEY: -
LOCATION: 412..414
OTHER INFORMATION: /note= "stop codon"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..414
US-08-256-261-13

alignment_scores:
Quality: 687.00 Length: 136
Ratio: 5.127 Gaps: 0
Percent Similarity: 98.529 Percent Identity: 96.324

alignment_block:
US-09-728-670-10 x US-08-256-261-13 ..
Align seg 1/1 to: US-08-256-261-13 from: 1 to: 414

1 SerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTyr 17
|||||
4 TCAAGTTCATTCGACAAAGGAAATATATAAAGCGATGACCGCGTTA 53
|||||
17 rphegiuProthGlyProTyrLeuMetValasnValThrGlyValAspS 34
|||||
54 TTTTGAACCAACAGCGCCGATTTGCTAGTAATGTGACTGAGAGTGAAG 103
|||||
34 erlysgLysangluLeuLeuSerProHisTyrValgluPheProIleLys 50
|||||

```

```

104 GTAAAGAAATGACTSYRTCCCKGKTATGTGAGTTTCTATTAA 153
51 ProgiuThrThreutThrLysGluLysIlegluTyrTyrValgluTPrAl 67
|||||
154 CTTGGAGCTACACTTACAAAAGAAAAATTAATCTATGTGCAATGGCC 203
|||||
67 aleuAspAlaThrAlaTyrLysGluPheArgValValgluLeuAspProS 84
|||||
204 ATTGATGTCGACAGCATATTAAGAGTTTAGAGTAGTTGAATTAGATCCA 253
|||||
84 erAlaLysIlegluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCCAAGATCGAAGTCCTTATATGATTAAGAAAGAAAAAGAGAA 303
|||||
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuS 117
|||||
304 ACGAAGTTTTCCTTATACAGAAAAAGTTTGTGTGCCAGATTATATC 353
|||||
117 rgluHisIleLysasnProgiuPheasnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAAAAACCTCGATTCACATTAAATTACMAAGTTGTATAG 403
|||||
134 luLysLys 136
|||||
404 AAAAGAAA 411

seq_name: /cgn2_6/prodata/1/lna/6A_COMB.seq:US-08-852-299-13
seq_documentation_block:
Sequence 13, Application US/08852299
Patent No. 6010897
GENERAL INFORMATION:
APPLICANT: Behnke, Detlef
APPLICANT: Schlott, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,299
FILING DATE: 17-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/256,261
FILING DATE:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "recombinant DNA"
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: phase 42D
IMMEDIATE SOURCE:
CLONE: Plasmid pMETS
FEATURE:

```

```

; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; US-08-852-299-13

```

```

alignment_scores:
  Quality: 687.00      Length: 136
  Ratio: 5.127        Gaps: 0
  Percent Similarity: 98.329   Percent Identity: 96.324

```

```

alignment_block:
  US-09-728-670-10 x US-08-852-299-13

```

```

Align seg 1/1 to: US-08-852-299-13 from: 1 to: 414

```

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTyr 17
  |||||
4 TCAGTTCATTCGACAAAGAAATATATAAAGCGATGCGGAGTTA 53
  |||||
17 rPhgIuPProThrGlyProTyrLeuMetValAsnValThgIlyValAsp 34
  |||||
54 TTTTGAACCAACAGCCCGATTGCTAGTAAATGTGACTGCGAGCTGATG 103
  |||||
34 eTlYsGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
  |||||
104 GTAAAGAAATGAGCTCTTTCCTCCCKGKRTGTCGAGTTCCATTTAA 153
  |||||
51 ProGluThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTyrPAl 67
  |||||
154 CCGTGGACTACACTTACAAAAGAAATTAATGATATGTCGAAATGGGC 203
  |||||
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
  |||||
204 ATTGATATCGCACACATATTAAGAGTTAGAGTGAATTAATGATCCAA 253
  |||||
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
  |||||
254 GCGCAAGATCGAAGTCACTATATGATTAAGATTAAGAAAAAGAAAGAA 303
  |||||
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
  |||||
304 ACGAAGCTCTTCCCTATTAACAGAAAAAGGTTTGTGTCCAGATTATTC 353
  |||||
117 rGluHisTleLysAsnProGlyPheAsnIleThrLysValValIleG 134
  |||||
354 AGAGCATATTAATAAACCTGGATTCACTTAATTACMAAGGTTGTATAG 403
  |||||
134 LuLysLys 136
  |||||
404 AAAAGAAA 411

```

```

seq_name: /cgn2_6/prodata/1/lna/5A.COMB.seq:US-08-256-261-7

```

```

seq_documentation_block:

```

```

; Sequence 7, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Helinz

```

```

; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; TITLE OF INVENTION: staphylokinases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256,261
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: phage 42d
; IMMEDIATE SOURCE:
; CLONE: Plasmid pMET5
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..384
; FEATURE:
; NAME/KEY: -
; LOCATION: 382..384
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..381
; OTHER INFORMATION: /product= "shortened mature
; OTHER INFORMATION: protein"
; US-08-256-261-7

```

```

alignment_scores:
  Quality: 643.00      Length: 126
  Ratio: 5.144        Gaps: 0
  Percent Similarity: 99.206   Percent Identity: 97.619

```

```

alignment_block:
  US-09-728-670-10 x US-08-256-261-7

```

```

Align seg 1/1 to: US-08-256-261-7 from: 1 to: 384

```

```

11 LysGlyAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVal 27
  |||||
4 AAAGCGATGACCGCGAGTTATTTGAACCAACAGGCCGTATTTGATGTT 53
  |||||
27 IAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisT 44
  |||||
54 AAATGTGACTGAGTGTGATGTAAGAAATGATGATTCCTCGCTT 103
  |||||
44 yValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60
  |||||
104 ATGTCAGTTTCTATTAACCTGGACTACACTTAACAAAAGAAATTT 153

```

```

61 GIUTYTYRVALGIUTRPALEUASPALATHTALATYRGLYSGIUPHEAR 77
154 GAATACATATGCGAATGGCATTTAGATCGACAGCATATAAGAGTTAG 203
77 gValValGluLeuAspProSerAlaLysIleGluValThrTyrTrpAspL 94
204 AGTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 253
94 yAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
254 AGAATTAAGAAAAAGAACAGAACGAGTCTTCCCTATACGAAAAAGCT 303
111 PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnLe 127
304 TTGTGTGCCGAGATTATACAGACATATTAATAAACCTGGATTCACTT 353
127 uIleThrLysValValIleGluLysLys 136
354 AATTACMAAGGTTGTTATAGAAAAAGAAA 381

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-852-299-7

seq_documentation_block:
; Sequence 7, Application US/08852299
; Patent No. 6010897
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (PRO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,299
; FILING DATE: 17-May-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,261
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phase 42D
; IMMEDIATE SOURCE:
; CLONE: Plasmid pMET5
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: 1..384
; FEATURE:
; NAME/KEY: -
; LOCATION: 382..384
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..381
; OTHER INFORMATION: /product= "shortened mature
; OTHER INFORMATION: protein"
; US-08-852-299-7

alignment_scores:
Quality: 643.00 Length: 126
Ratio: 5.144 Gaps: 0
Percent Similarity: 99.206 Percent Identity: 97.619

alignment_block:
US-09-728-670-10 x US-08-852-299-7 ..

Align seg 1/1 to: US-08-852-299-7 from: 1 to: 384

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuKetVa 27
4 AAAGCGATGACGCGAGTATTTTGAACCAACAGCCCGTATTGATGGT 53
27 lAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisT 44
54 AAATGCTGCTGAGTGTATGATAAAGAAATGCTATGCCCTCGT 103
44 yValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60
104 ATGCGAGTTTCCATTAACCTGCGACTACACTTAAAAAGAAAAAATT 153
61 GIUTYTYRVALGIUTRPALEUASPALATHTALATYRGLYSGIUPHEAR 77
154 GAATACATATGCGAATGGCATTTAGATCGACAGCATATAAGAGTTAG 203
77 gValValGluLeuAspProSerAlaLysIleGluValThrTyrTrpAspL 94
204 AGTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 253
94 yAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
254 AGAATTAAGAAAAAGAACAGAACGAGTCTTCCCTATACGAAAAAGCT 303
111 PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnLe 127
304 TTGTGTGCCGAGATTATACAGACATATTAATAAACCTGGATTCACTT 353
127 uIleThrLysValValIleGluLysLys 136
354 AATTACMAAGGTTGTTATAGAAAAAGAAA 381

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-9

seq_documentation_block:
; Sequence 9, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlotz, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York

```

```

STATE: NEW YORK
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "recombinant DNA"
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Phase 4ZD
IMMEDIATE SOURCE:
CLONE: plasmid pMER5
FEATURE:
NAME/KEY: -
LOCATION: 1..3
OTHER INFORMATION: /note= "start codon"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..372
FEATURE:
NAME/KEY: -
LOCATION: 370..372
OTHER INFORMATION: /note= "stop codon"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 4..369
OTHER INFORMATION: /product= "shortened mature protein"
OTHER INFORMATION: protein"
US-08-256-261-9

alignment_scores:
Quality: 620.00 Length: 122
Ratio: 5.124 Gaps: 0
Percent Similarity: 99.180 Percent Identity: 97.541

alignment_block:
US-09-728-670-10 x US-08-256-261-9 ..

Align seg 1/1 to: US-08-256-261-9 from: 1 to: 372

15 AAlasErTyRpheGluProThrGlYPToTyRLeuMeTValnsnValThrGI 31
|||||
4 GCgAGtATTTTGAAcCAACAAGCCCGGTATTGGATGGAATGGACTGG 53
|||||
31 yAlAsPseerTySGlYaNGUlueLseRPrOhIstYrYalGlubher 48
|||||
54 AcRTAgtAGGTAAAGAAMATGAATGCTATTCGCCCTGTATGTcGAgTTTC 103
|||||
48 rOllElAsPrOcLIgYhThrLeuFrTrLySGlULysIIeGLuTYrTYVAl 64
|||||
104 cHAtTAACAcTcGCACtACtACTACCAAAGAAAAATTGAATAcTAcTc 153
|||||
65 GlUTTPAlAleuAsPAlThrAlATyrLySGlUbheArGaYAlValGIule 81
|||||
154 GAATGGCGCTTGGAGTCACACAGCATATGAAGAGTTTAGAGTAcTTGAAT 203
|||||
81 uAsPProSeRalALysIIeGLUvAlThrTYrTYAsPlYASnLYSL 98
|||||
204 AcATCcAAAGCGAAAGATCGAGTAcTATTAATGATAGAAATAAGAAA 253

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

OTHER INFORMATION: /product= "shortened mature
OTHER INFORMATION: protein"
US-08-852-299-9

alignment_scores:
Quality: 620.00 Length: 122
Ratio: 5.124 Gaps: 0
Percent Similarity: 99.180 Percent Identity: 97.541

alignment_block:
US-09-728-670-10 x US-08-852-299-9 ..

Align seg 1/1 to: US-08-852-299-9 from: 1 to: 372

```

15 AlaSerTyPheGluProThrGlyProTyrLeuMetValAsnValThrG1
|||||
4 GCGAGTATTGTTGAACCAACAGCCCGTATTGATGCTAAATGTGACTGG 53
31 yValAspSerLysGlyAsnGluLeuLeuSerProHisTyrValGluPheP 48
|||||
54 AGTTGATGGTAAAGAAATGAATGCTATCCCTCGTATGTCGAGTTTC 103
48 roLeLysProGlyThrThrLeuThrLysGluLysIleGluTyrTyrVal 64
|||||
104 CTATTAAACCTGGACTACACTTACAAAAGAAATTAATGAAATCATATGTC 153
65 GluThrPalaLeuAspAlaThrAlaTyrLysGluPheArgValAlaGluDe 81
|||||
154 GAATGGCGATTAGATCCGACACATATAAGAGTTAGAGTACTGTAATT 203
81 uAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnLysLysL 98
|||||
204 AGATCCAGACGCAAGATCGAAGTCACTTATATGATTAAGATAAGAAA 253
98 ySgluGluThrLysSerPheProIleThrGluLysGlyPheValValPro 114
|||||
254 AAGAAAGAAAGCAAGCTCTTCCCTATACAGAAAAGGTTTGTGCCCA 303
115 AspLeuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysVa 131
304 GATTATTCAGACATATTAAAAACCTGATTCACACTTAATTACAAAGCT 353
131 lValIleGluLysLys 136
|||||
354 TGTATATAGAAAAGAAA 369

```

seq_name: /cgn2_6/ptodata/1/lna/backfile1.seq:5182210-9

seq_documentation_block:

Patent No. 5182210
APPLICANT: BINNS, MATTHEW M.;BOURSNELL, MICHAEL E.G.;
CAMPBELL, JOAN I.A.;TOMLEY, FIONA M.
TITLE OF INVENTION: FOWLPOX VIRUS PROMOTERS
NUMBER OF SEQUENCES: 22
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,608
FILING DATE: 21-OCT-1988
SEQ ID NO:9;
LENGTH: 11225
5182210-9

alignment_scores:

Quality: 84.50 Length: 123
Ratio: 1.408 Gaps: 5
Percent Similarity: 48.780 Percent Identity: 25.203

alignment_block:

US-09-728-670-10 x 5182210-9/rev ..

Align seg 1/1 to reverse of: 5182210-9 from: 1 to: 11225

```

43 HisTyrValGluPheProIleLysProGlyThrThrLeuThrLysGluLys 59
|||||
1113 CATATATGTTGAAACCTATATACATATAGCTGCTAATAGAACAGAAAGTAA 11064
59 sIle...GluTyrTyrValGluThrPalaLeuAspAlaThrAla..... 72
|||||
11063 AATAGTAAATTTGCTTATAGATACGGCGCTGACATTAAATTCAGAAAGACG 11014
73 .....TyrLysGlu 75
11013 GCGCGAATGCTAATACCTATACATACGCTATAGAAAGTTACGATCCG 10964
76 PheArgVal.....ValGluLeuAspProSerAlaLysIleG1 88
|||||
10963 TTATGATTAATAAATAATTAAGTATTTATAGACCAAGCGGCC..... 10922
88 uValThrTyrTyrAspLysAsnLysLysGluGluThrLysSerPheP 105
|||||
10921 .....GATATTAACAAACAAAGCGTTTAACTAATACATCC 10885
105 roLeThrGluLysGlyPheValValProAspLeuSerGluHisIleLys 121
|||||
10884 CCTTATACGAACACTAGCTTTATTTACCGACGACCTATTAGATTACATCAT 10835
122 AsnProGlyPheAsnLeu.....IleTh 129
10834 TCTAGAGGAGCTAATATTAATATAAAGACAGATGGTAGAATATATTAC 10785
129 rLysValValIleGluLys 135
|||||
10784 ACGAATTAATATTAGAACGA 10766

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-09-321-831-1

seq_documentation_block:

Sequence 1, Application US/09321831
Patent No. 6262336
GENERAL INFORMATION:
APPLICANT: LUBON, Henryk
APPLICANT: DROHAN, William
APPLICANT: HENNIGHAUSEN, Lothar
APPLICANT: VELANDER, William
TITLE OF INVENTION: Expression of a Heterologous Protein C in Mammary Tissue of
FILE REFERENCE: 030523/0150
CURRENT APPLICATION NUMBER: US/09/321,831
CURRENT FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: US 09/184,163
PRIOR FILING DATE: 1998-11-02
PRIOR APPLICATION NUMBER: US 07/943,246
PRIOR FILING DATE: 1992-09-10
PRIOR APPLICATION NUMBER: US 07/638,995
PRIOR FILING DATE: 1991-01-11
PRIOR APPLICATION NUMBER: PCT/US98/02638
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 08/443,184
PRIOR FILING DATE: 1995-05-17
PRIOR APPLICATION NUMBER: US 08/198,068
PRIOR FILING DATE: 1994-02-18
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 4122
TYPE: DNA
ORGANISM: WAP gene promoter fragment from the C57B/6 mouse strain
FEATURE:
NAME/KEY: misc_feature
LOCATION: (544)..(544)
OTHER INFORMATION: "n" is unknown
NAME/KEY: misc_feature
LOCATION: (3377)..(544)
OTHER INFORMATION: "n" is unknown
US-09-321-831-1

APPLICANT: Dunn, Martha

; APPLICANT: Chen, Jeng S.


```

65 .....glutpalaLeuaspalaThrAla..... 72
187 TATGCAAGCATTCATGACAAATGTCATGACAAAGTGAAGG 236
73 .....TyrLysGluPheArgValAlaGluLeuAspProSerAlaLysI 87
237 ATGATTTATGAGAAATA.....TATTCAGACTCCTCCAGAA 274
87 leglu.....ValThrTyrTyrAspLysAsnLysLysGluGlu 100
275 TTCAACCTTGATGAACCATATAACTATATCTGAAAAAGAAAAAGAA 324
101 ThrLysSerPheProIleThrGlu 108
325 ATTTATCGCTGCCCAATCAAGAA 348

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-480-481-4

seq_documentation_block:
; Sequence 4, Application US/08480481
; Patent No. 5639614
; GENERAL INFORMATION:
; APPLICANT: Sen, Luyi
; APPLICANT: Phillips, Kenneth D.
; TITLE OF INVENTION: GENE MUTATION IN PATIENTS WITH IDIOPATHIC DILATED CARD
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,481
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: UCLAO12.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1050 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Other
; LOCATION: 380...380
; OTHER INFORMATION: Position distinguishes idiopathic
; OTHER INFORMATION: dilated cardiomyopathy
; NAME/KEY: Other
; LOCATION: 776...776

```

```

; OTHER INFORMATION: Position distinguishes idiopathic
; OTHER INFORMATION: dilated cardiomyopathy
; NAME/KEY: Other
; LOCATION: 1...1
; OTHER INFORMATION: Position 1 corresponds to
; OTHER INFORMATION: 7987 of the rabbit cDNA homolog
; US-08-480-481-4

alignment_scores:
Quality: 73.00 Length: 108
Ratio: 1.281 Gaps: 6
Percent Similarity: 52.778 Percent Identity: 25.000

alignment_block:
US-09-728-670-10 x US-08-480-481-4 ..

Align seg 1/1 to: US-08-480-481-4 from: 1 to: 1050

17 TyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAs 33
|||||...|||...|||...|||...|||...|||...|||...|||...|||
55 TACATGAAATCAAAATTATGTCAGTATGATGGAAGAAACATCATCAATGGA 104
33 pSerLysGluAsnGluLeuLeuSerProHisTyrValGluPheProIleL 50
|||||...|||...|||...|||...|||...|||...|||...|||...|||
105 TTCGAAAGGAAAC.....TTTAAACCAAA.....CCTGTG 136
50 ySPrgLlYThrThrLeuThrLysGluLysIleGluTyrTyrVal..... 64
|||||...|||...|||...|||...|||...|||...|||...|||...|||
137 ATACCTCAAAATTTATATCTCTGAGAAAGTTGGAATCTTATTAACAA 186
65 .....glutpalaLeuaspalaThrAla..... 72
187 TATGCAAGCATTCATGACAAATGTCATGACAAAGTGAAGG 236
73 .....TyrLysGluPheArgValAlaGluLeuAspProSerAlaLysI 87
237 ATGATTTATGAGAAATA.....TATTCAGACTCCTCCAGAA 274
87 leglu.....ValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||...|||...|||...|||...|||...|||...|||...|||...|||
275 TTCAACCTTGATGAACCATATAACTATATCTGAAAAAGAAAAAGAA 324
101 ThrLysSerPheProIleThrGlu 108
325 ATTTATCGCTGCCCAATCAAGAA 348

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-525-654A-2

seq_documentation_block:
; Sequence 2, Application US/08525654A
; Patent No. 5736356
; GENERAL INFORMATION:
; APPLICANT: SANO, KOICHIRO
; APPLICANT: KIMAZAWA, YOSHIYUKI
; APPLICANT: YASEUDA, HISASHI
; APPLICANT: SEGURO, KATSUYA
; APPLICANT: MOTOKI, MASAO
; TITLE OF INVENTION: TRANSGLUTAMINASE ORIGINATED FROM
; NUMBER OF SEQUENCES: 150
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSER: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

APPLICANT: SEGURO, KATSUYA

[illegible]

2209 CCGTGAAGTATGACGACCTTCACCTCCAAACAGATTGTCGACATTGACAC 2258

83 oSerAlaIysAlIeGluValIhTyrIAspIysAsnIysIysGlu 99
 2259 CACGCGCGACATTAGGTGATCCGACACAAACAAATTAACGACGAC 2307

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-469-005A-9

seq_documentation_block:

; Sequence 9, Application US/08469005A

; Patent No. 5665874

; GENERAL INFORMATION:

; APPLICANT: KUNDA, FRANCIS P.

; APPLICANT: PASTERNAK, GARY A.

; TITLE OF INVENTION: CANCER RELATED ANTIGEN

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS: 14

; ADDRESSEE: BAKER & BOTS, L.L.P.

; STREET: 1299 Pennsylvania Avenue, N.W.

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20004-2400

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/469,005A

; FILING DATE: 05-JUN-1995

; CLASSIFICATION: 536

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 08/188,426

; FILING DATE: 24-JAN-1994

; APPLICATION NUMBER: 08/096,908

; FILING DATE: 26-JUL-1993

; APPLICATION NUMBER: 07/917,716

; FILING DATE: 24-JUL-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Posorske, Laurence H

; REGISTRATION NUMBER: 34,698

; REFERENCE/DOCKET NUMBER: 062482-0113

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-639-7700

; TELEFAX: 202-639-7890

; TELEX:

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8460 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE:

; ORIGINAL SOURCE:

; FEATURE:

; NAME/KEY: Coding Sequence

; LOCATION: 124...7650

; OTHER INFORMATION:

; US-08-469-005A-9

alignment_scores:

Quality: 73.00 Length: 61

Ratio: 1.921 Gaps: 1

Percent Similarity: 62.295 Percent Identity: 34.426

Alignment_block:

US-09-728-670-10 x US-08-469-005A-9 ..

Align seg 1/1 to: US-08-469-005A-9 from: 1 to: 8460

27 ValAsnValIhThGlyValAspSerIysGlyAsnGluLeuSerProH 43
 2530 CTGACACCTCTGAGCATCGACGCCAACCCAAATGCTGTGCCACCT.. 2577

43 sTyrValGluPheProIleIysProGlyThrThrLeuThrIysGluIys 60

2578 ...GTGAGTTCGCCAGCTCCCGAGAACCTCCCTCATCTCCCACTCA 2623

60 IeGluTyrTyrValGluTrrPalalaLeuAspAlaThrAlaTyrIysGluPhe 76

2624 TCAAGTGGACCAACAGCCTGGCTGGACGTCCGCGCCGCGGAGACTTC 2673

77 ArgValValGluLeuAspProSerAlaIysIle 87

2674 CCCAAGGTTCAAGTTCCTCCCTCAGCCGCAATC 2706

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-261-907-1

seq_documentation_block:

; Sequence 1, Application US/09261907A

; Patent No. 6294364

; GENERAL INFORMATION:

; APPLICANT: ELLIS, CATHERINE

; APPLICANT: LONSDALE, JOHN

; APPLICANT: BERGSMAN, DEK J.

; APPLICANT: MOONEY, JEFFREY L.

; APPLICANT: DEPIERA, MEGAN E.

; APPLICANT: CHAPMAN, CONRAD

; TITLE OF INVENTION: HUMAN FAS

; FILE REFERENCE: GP-70603

; CURRENT APPLICATION NUMBER: US/09/261,907A

; CURRENT FILING DATE: 1999-03-03

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 8519

; TYPE: DNA

; ORGANISM: HOMO SAPIENS

; US-09-261-907-1

alignment_scores:

Quality: 73.00 Length: 61

Ratio: 1.921 Gaps: 1

Percent Similarity: 62.295 Percent Identity: 34.426

Alignment_block:

US-09-728-670-10 x US-09-261-907-1 ..

Align seg 1/1 to: US-09-261-907-1 from: 1 to: 8519

27 ValAsnValIhThGlyValAspSerIysGlyAsnGluLeuSerProH 43

2562 CTGACACCTCTGAGCATCGACGCCAACCCAAATGCTGTGCCACCT.. 2609

43 sTyrValGluPheProIleIysProGlyThrThrLeuThrIysGluIys 60

2610 ...GTGAGTTCGCCAGCTCCCGAGAACCTCCCTCATCTCCCACTCA 2655

60 IeGluTyrTyrValGluTrrPalalaLeuAspAlaThrAlaTyrIysGluPhe 76

2656 TCAAGTGGACCAACAGCCTGGCTGGACGTCCGCGCCGCGGAGACTTC 2705

77 ArgValValGluLeuAspProSerAlaIysIle 87

2706 CCCAAGGTTCAAGTTCCTCCCTCAGCCGCAATC 2738

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-282-305-9

seq_documentation_block:

; Sequence 9, Application US/09282305


```

seq_name: /cgn2_6/ploddata/1/lna/5B_COMB.seq:US-08-795-475-4
590 ACGAAGACGGCTTTTCATTTGGTT 612

seq_documentation_block:
Sequence 4, Application US/08795475
Patent No. 5965390
GENERAL INFORMATION:
APPLICANT: Bjvck, Lars
APPLICANT: Sjvck, Ulf
TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,475
FILING DATE: 11-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McMASTERS, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 100084,402D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1308 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli U392/PHDLG, DSM 7055
US-08-795-475-4

alignment_scores:
Quality: 69.00 Length: 116
Ratio: 1.190 Gaps: 6
Percent Similarity: 50.000 Percent Identity: 31.034

alignment_block:
Align seg 1/1 to: US-08-795-475-4 from: 1 to: 1308
US-09-728-670-10 x US-08-795-475-4 ..

5 AspLySGlYtStYrLyGlyAspAspLaseTyrPheGluProth 21
:::|||||
823 GAAATTCGTAATAT.....ACACAGACTTGAAGATCG 857
21 rGlyProTyrLeuMetValAsnValThrGly.....ValAspSerLysG 36
||| :::::|||||
858 TCGATACACTATATATTAGATTGTCAGTAAAGATTGACGAAAAAC 907
36 LyAsnGluLeuLeuSerProHisTyrValGluGluPheProIleLysProGly 52
:::||||| ::::: |||
908 CAGAAGAACCCCTGGACACTTAACAATTA.....ATCCTTAATGCT 948
53 ThrThrLeuThrLySGlYsIleGluTyrTyrValGluTrpAlaLeuAs 69
||||| |||
949 AAAACATTGAAGGCGGAACAACACTACGACCTGTTGAT.....GC 989

```

```

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-064-693A-22
103 SerpHeProIleThrGlyLysGlyPheValValProAspLeuSerGlu 118
1072 ACCTTACAGTACTGCAAAAACGAGAGTATC...GATGCGTGTGAA 1116
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-064-693A-22
seq_documentation_block:
; Sequence 22, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INEL-LOCKHEED Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1348 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-064-693A-22
alignment_scores:
Quality: 69.00 Length: 106
Ratio: 1.302 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 24.528
alignment_block:
US-09-728-670-10 x US-09-064-693A-22 ..
Align seg 1/1 to: US-09-064-693A-22 from: 1 to: 1348
11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27
|||||
:::

```


PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/029,848
 FILING DATE: 30-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Saliwanchik, David R.
 REGISTRATION NUMBER: 31,794
 REFERENCE/DOCKET NUMBER: MA-708
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 352-375-8100
 TELEFAX: 352-372-5800
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1278 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: 68F
 US-09-960-780-26

alignment_scores:
 Quality: 68.50 Length: 153
 Ratio: 0.867 Gaps: 8
 Percent Similarity: 51.634 Percent Identity: 20.261

alignment_block:
 US-09-728-670-10 x US-09-960-780-26 ..

Align seg 1/1 to: US-09-960-780-26 from: 1 to: 1278

```

10 LysLysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeu 26
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 AAGAGAACGACGCGATTTACATTAACTTATCAGATGATGACATACGAT 216
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
26 tvalaenvalthrGly.....ValAspSerLysGlyAsnGluLeu 41
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 GATAGAAATGATGATGGAACTTATTTCTAATTAAGGAAAGAAACAC 266
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41 eproniSTyrValGluPheProLleLysProGlyThrThrLeuThrLys 57
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
267 TTGTCCAT.....TTAGAAAAGGA...CAGTTCGTTTC 298
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58 GluLysLleGluTyrTyrValGlu.....TrrAlaLeuAspAlaThrAl 72
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
299 ATCAAAATTAATATTCAACTGATGAAACCATTTAATGCGGATGATCA 348
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
72 aTyLysGluPheArGValValGluLeuAspProSerAlaLysLleGlu 89
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
349 CTTTAAATAATTTGAACCTCTTAAAGTAGTACTAAGCAACAGTCCAGC 398
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
89 alThrTyrTyrAspLys.....AsnLysLysGlu 99
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
399 AATATTCAGTAGATGATTAAGAAACCCGATTTAATAAAGAAAGCA 448
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
100 Glu.....ThrLysSerPheProLleThrGluLys.. 109
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
449 CAAGAAATTTTAACAAAGCAACAAACAAACCTTATTACTCAAAAGT 498
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
110 .....GlyPheValValPro 115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
499 GAAGAGTACTAGGATGAAGACACGAGTACAGATTCGATTCATTCAC 548
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
115 spLeuSerGluHisLleLysAsnProGlyPheAsnLeuLleThrLysVal 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
549 ACATTTGGGAA.....GAAGATGGTATACATCATCAAAATTAAGAT 589
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
132 ValLleGlu 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
590 GCCGTCAAA 598
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: /cgn2_6/pdata/ta/1/lna/6B_COMB.seq: US-09-073-898-26

seq_documentation_block:
 Sequence 26, Application US/09073898
 Patent No. 6242669
 GENERAL INFORMATION:
 APPLICANT: Feltelson, Jerald S.
 APPLICANT: Schnepf, H. Ernest
 APPLICANT: Narva, Kenneth E.
 APPLICANT: Stockhoff, Brian A.
 APPLICANT: Schmeltz, James
 APPLICANT: Loewer, David
 APPLICANT: Dullum, Charles Joseph
 APPLICANT: Muller-Cohn, Judy
 APPLICANT: Stamp, Lisa
 APPLICANT: Morrill, George
 APPLICANT: Finstad-Lee, Stacey
 TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
 TITLE OF INVENTION: Sequences Which Encode These Toxins
 NUMBER OF SEQUENCES: 144
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
 STREET: 2421 N.W. 41st Street, Suite A-1
 CITY: Gainesville
 STATE: FL
 COUNTRY: US
 ZIP: 32606-6669
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/073,898
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/029,848
 FILING DATE: 30-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/960,780
 FILING DATE: 30-OCT-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Sanders, Jay M.
 REGISTRATION NUMBER: 39,355
 REFERENCE/DOCKET NUMBER: MA-708C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 352-375-8100
 TELEFAX: 352-372-5800
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1278 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: 68F
 US-09-073-898-26

alignment_scores:
 Quality: 68.50 Length: 153
 Ratio: 0.867 Gaps: 8
 Percent Similarity: 51.634 Percent Identity: 20.261

alignment_block:
 US-09-728-670-10 x US-09-073-898-26 ..

Align seg 1/1 to: US-09-073-898-26 from: 1 to: 1278

```

10 LysLysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeu 26
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 AAGAGAACGACGCGATTTACATTAACTTATCAGATGATGACATACGAT 216

```



```

26 tValAsnValhrcglu.....ValAspSerLysGlyAsnGluLeuLeus 41
27 |-----|
28 |-----|
29 ATCAAAAGATAGATATCAAGCTGACGAACCACTTATCGGATAGTCAAC 348
72 aTyriScgluPhcArgValValGluLeuAspProSerAlaLysIleGlu 89
349 CTTTAAATTTGAAACCTTTTAAAGTATGATCTTACCAACACGCCAGC 358
89 alhTrTyTAspLys.....AsnLysLysGlu 99
399 AAATTCACACTGATGATATTAAGAAACCGTGAATTTATTAAGAAACA 448
100 Glu.....ThlySerPhePheIleThrcgluLys.. 109
449 CAAGATTTCTTACAAAGACAAACAAACAACTTTTACTCAAAAAGT 498
110 .....GlyPheValValPro 115
499 GAAGGACTGATGAGTGAAGACACGAGTACAGATGAGATTTCTATTCAG 548
115 sPLeSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysVal 131
549 ACATTTGGGAA.....GAAATGGGTATCCATTCACAAATAAAGATT 588
132 ValIleGlu 134
590 GCCGTCAAA 598

seq_name: /cgn2_6/prodata/1/lna/5A_COMB.seq:US-08-257-073-10

seq documentation block:
Sequence 10: Application US/08257073
Patent No. 576597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtiss, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
City: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 5181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-10

Alignment_scores:
Quality: 66.50 Length: 103
Ratio: 1.245 Gaps: 6
Percent Similarity: 53.398 Percent Identity: 30.097

alignment_block:
US-09-728-670-10 x US-08-257-073-10 ..

Align seg 1/1 to: US-08-257-073-10 from: 1 to: 5181

28 AsnValThrGlyValAlaSpSerIysGlyAsnGluLeuLeuSerProHisTy4
||||| : : : : : ||||| : : : : :
1021 AACATTAAGGAATTAATCTGATAGATTAAGATTAAGAAATTTAAATCCCCACC 1070
44 rValGluPheProIleLysProGlyThrThrLeuThrLys...GluLysI 60
: : : : : ||| : : : : :
1071 GGCCCAATTCGGAATAATACACCAATATCTCTCTGATTAAGACACAAAAAAA 1120
60 leGluThrTyValValIgluThrAlaLeuAspAlaThrAlaTyLysGluPhe 76
||||| : : : : : ||| : : : : :
1121 TCGAGNACACGAGAAAAATAAAGAAATTGCCAAACTATTAAATTT 1170
77 ArgValValGluLeu.....AspProSerAlaLysIleGluValThrTy 91
: : : : : ||||| : : : : :
1171 AACATGATATGTTATTATTACTGATCA.....CTGAAATTAAGATA 1211
91 rTyT.....AspLysAsnLysLys.....LysG 99
||||| : : : : : ||||| : : : : :
1212 TTATTTAAGAGAAAAATAAATAAAAGTTGATGTAAACCTAAATCACAAAG 1261
99 IuGluThrLysSerPheProIleThrGlu.....LysGlyPhe 111
: : : : : ||||| : : : : :
1262 ATCTTCAGAAATCTGTTCAATATCCAAATAGTCTTATCCAAATGTATT 1311
112 ValValPro 114
||| |||
1312 GTATATCTCT 1320

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-198-956-1

seq_documentation_block:
Sequence 1, Application US/09198956
Patent No. 6165769
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schuelein, Martin
APPLICANT: Lange, Niels Erik K.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Schnoor, Kirk
TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
FILE REFERENCE: 5377.200-US
CURRENT FILING DATE: 1998-11-24
EARLIER APPLICATION NUMBER: US/09/198,956
EARLIER FILING DATE: 1997-11-24

```

```
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1485
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-198-956-1

alignment_scores:
  Quality: 68.00      Length: 113
  Ratio: 1.193      Gaps: 5
  Percent Similarity: 50.442      Percent Identity: 23.894

alignment_block:
US-09-728-670-10 x US-09-198-956-1  ..

Align seg 1/1 to: US-09-198-956-1 from: 1 to: 1485

16 SerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyVal 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 TCTTTTGGCTGCGCGCGCATTTATTTTATAGCT...ACAGTCCC 80
32 LAspSerLySGlyAsnGluLeuLeuSerProHisTyrValGluPhePro 49
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 TCGTTCTGCGCAACGAAAGATTATCCGAAACAGATGATCAGCTGGAAA 130
49 LeysProGlyThrThrLeuThrLys..... 57
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 GCTCATCAGCTTAAATATTAATACGCCCGCGCAATCAAGCAACGCCCG 180
58 .....GluLysIleGluTyrTyrValGluTr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 TTAACAGCAAAACAGACGCGGAGAAAAGAA.....AGATG 221
66 PalaleuAspAlaThrAlaTyrLySGluPheArgValGluLeuAsp 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 GAGGCTTGATACGTCTGACGCAAAACATTCAAATCAGAAATATGAT. 270
83 roSerAlaLysIle..... 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 ..AGCGCAAAATATATCATCCCTGCCATTCACGACGTGCAGACAATAT 318
319 CCGGCTGTGCTACTATGACAAATTCACGAGAGAGAG 357

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-131-648-4

seq_documentation_block:
; Sequence 4, Application US/09131648
; Patent No. 6188920
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guebler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
; FILE REFERENCE: PF-0576 US
; CURRENT APPLICATION NUMBER: US/09/131,648
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 2290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2687731
US-09-131-648-4
```

```
alignment_scores:
  Quality: 67.00      Length: 106
  Ratio: 1.098      Gaps: 5
  Percent Similarity: 57.547      Percent Identity: 25.472

alignment_block:
US-09-728-670-10 x US-09-131-648-4  ..

Align seg 1/1 to: US-09-131-648-4 from: 1 to: 2290

27 ValAsnValThrGlyValAspSerLySGlyAsnGluLeu..... 40
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1635 TTGAATATTAATAATAGAAATATTCAGCGCCAAATTCAGTTGGTCTCG 1684
41 .....SerProHisTyrValGluPheProIleLysProGlyThrL 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1685 GAAGCAAGTTCTAAATTCACAAATCTAGTTAA...TGCACAGCT 1731
55 eutThrLySGlyIleGluTyrTyrValGluTrPalaleuAspAlaThr 71
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1732 TTGCAAGACTGAATAATTCATGCTGCGCAAGTGCGAATACATCT 1781
72 AlaTyrLySGluPheArgValGluLeuAspProSerAlaLysIleG 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1782 GATGTCAAGGTATTAATCTACTCATCTGAAATCCATCACTGATATA 1831
88 uVal.....ThrTyrTyrAspLysAsnLysLysLysGlu 100
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1832 AATTTGATGATATTCCTCCACCACTATCAGAAACAGAAA..... 1874
100 LuThrLySerPheProIleThrGluLySGlyPheValProAspLeu 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1875 .....AATGTGTAATGTCCACCAAAAGT...TTGCACCTGATCA 1916
117 SerLuHisIleLysAsn 122
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1917 AAGAGTATGAAGAAAT 1934

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-968-563-7

seq_documentation_block:
; Sequence 7, Application US/08968563
; Patent No. 6013494
; GENERAL INFORMATION:
; APPLICANT: CHARLES E. NAKAMURA
; APPLICANT: ANTHONY A. GATENBY
; APPLICANT: AMY (KUANG-HUA) HSU
; APPLICANT: RICHARD D. LA REAU
; APPLICANT: SHARON L. HAYNE
; APPLICANT: MARIA DIAZ-TORRES
; APPLICANT: DONALD E. TRIMBUR
; APPLICANT: GREGORY M. WHITED
; APPLICANT: VASANTHA NAGARAJAN
; APPLICANT: MARK S. PAYNE
; APPLICANT: STEPHEN K. PICATAGGIO
; APPLICANT: RAMESH V. NAIR
; TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
; TITLE OF INVENTION: PRODUCTION OF 1,3-PROPANEDIOL
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 4 CAMBRIDGE PLACE
; STREET: 1870 SOUTH WINTON ROAD
; CITY: ROCHESTER
; STATE: NEW YORK
```

```

? COUNTRY: U.S.A.
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.50 INCH DISKETTE
? OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
? SOFTWARE: MICROSOFT WORD VERSION 7.0A
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/968,563
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 60/030,601
? FILING DATE: NOVEMBER 13, 1996
? ATTORNEY/AGENT INFORMATION:
? NAME: FLOYD, LINDA AXAMETHY
? REGISTRATION NUMBER: 33,692
? REFERENCE/DOCKET NUMBER: CR-9982
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 302-892-8112
? TELEFAX: 302-773-0164
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3178 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: GUT2
? US-08-968-563-7

```

```

alignment_scores:
  Quality: 67.00      Length: 112
  Ratio: 1.063        Gaps: 4
  Percent Similarity: 56.250      Percent Identity: 23.214

```

```

align seg 1/1 to: US-08-968-563-7 ..
US-09-728-670-10 x US-08-968-563-7

```

```

29 ValThrglyValAspSerlySglnGluLeuSerProHisTyrVa 45
   ::::::::::::::::::::
1053 ATCAGCGGTGTGGAGAACGGCCCTACCGCTTGATC.....TATGT 1093
45 lGluPhePro.....lIeLysProGlyThrThr..... 54
   |||||
1094 CGAGGTACAAATAATGATCAAGACCCCACTTCTGTAGATTGCGGTG 1143
55 .....LeuThrLysGluLyslIeGluTyrTyrValGlu 65
   |||||
1144 CGAGGCGCGGGAGCTTGAGACTAATGAGCTGTGCAGATCAACGCTAA 1193
66 TrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValAlGluLeu 82
   ::::::::::::::::::::
1194 TGTGTGTGTAATGCGGCGGCGCATACAGTGAAGCCATTTCATGAA 1243
82 P.....ProSerAlaLyslIeGluValThrTyrTrpAspLysAln 97
   |||||
1244 CGCAACCCATCCGGTCTGCCGAGCTCCCGCTAAACGACAACTCCAA 1293
97 yalysGluLgluThrLysSerPheProIleThrGluLysGlyPheVal 113
   |||||
1294 TCAGAGTCGACTTTCATCAATCTCGTCATGAGCCCGAATAATGTG 1343
114 TrpAspLeuSerGluHislIeLysAsnProGlyPhe 125
   |||||
1344 CCATCTATTGGGCTTCATCGTATGCGCTCTTT 1379

```

```

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-969-683A-7
seq_documentation_block:

```

```

? Sequence 7, Application US/08969683A
? Patent No. 6136576
? GENERAL INFORMATION:
? APPLICANT: GENENCOR INTERNATIONAL, INC.
? TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
? TITLE OF INVENTION: PRODUCTION OF 1,3 PROPANEDIOL
? NUMBER OF SEQUENCES: 68
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genencor International, Inc.
? STREET: 4 Cambridge Place
? STREET: 1870 South Winton Road
? CITY: Rochester
? STATE: NY
? COUNTRY: U.S.A
? ZIP: 14618
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows
? SOFTWARE: FASTSEQ for Windows Version 2.0b
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/969,683A
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/US97/20873
? FILING DATE: 13-NOV-1997
? APPLICATION NUMBER: 60/030,601
? FILING DATE: 13-NOV-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Galister, Debra
? REGISTRATION NUMBER: 33,888
? REFERENCE/DOCKET NUMBER: GC 369-2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650-864-7620
? TELEFAX: 650-845-6504
? TELEX:
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3178 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: GUT2
? US-08-969-683A-7

```

```

alignment_scores:
  Quality: 67.00      Length: 112
  Ratio: 1.063        Gaps: 4
  Percent Similarity: 56.250      Percent Identity: 23.214

```

```

align seg 1/1 to: US-08-969-683A-7 ..
US-09-728-670-10 x US-08-969-683A-7

```

```

29 ValThrglyValAspSerlySglnGluLeuSerProHisTyrVa 45
   ::::::::::::::::::::
1053 ATCAGCGGTGTGGAGAACGGCGCTACCGCTTGATC.....TATGT 1093
45 lGluPhePro.....lIeLysProGlyThrThr..... 54
   |||||
1094 CGAGGTACAAATAATGATCAAGACCCCACTTCTGTAGATTGCGGTG 1143
55 .....LeuThrLysGluLyslIeGluTyrTyrValGlu 65
   |||||
1144 CGAGGCGCGGGAGCTTGAGACTAATGAGCTGTGCAGATCAACGCTAA 1193
66 TrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValAlGluLeu 82
   ::::::::::::::::::::

```

```

alignment_block:
US-09-728-670-10 x US-09-297-928-3 ..

Align seg 1/1 to: US-09-297-928-3 from: 1 to: 3178

29 valtrhrglyValAspserLysGLyAsnGluLeuLeuSerProHisrYrVa 45
|||||
1053 ATCAGCGGTGTGGAGAACGGCGCTACCGCTTGCATC.....TATGCT 1093

45 lgluPhePro.....lLeLysProGlyThr..... 54
|||||
1094 CGAGGTACAAAATTGATCAAGAACCCACACTCTCGTAAGCTATGCGTG 1143

55 .....LeuThrLysLysLysLysGluTyrTyrValLglu 65
|||||
1144 CCGAGCGCCGGGACGTGAGACTATATAGGTTGTCGATCAACACGCTAA 1193

66 TrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValAlaGluLeuAs 82
|||||
1194 TGTGTGTGTCATGCCACGCGCCCATACAGTACAGCCCATTTGGCAATGCA 1243

82 P.....ProSerAlaLysLysLysLysValThrTyrTyrAspLysAsnLysL 97
|||||
1244 CCGCAACCCATCGCGTGTGCCGAGCTCCCGCGTAACGACAACTCCAAAG 1293

97 yslYsgLysGluThrLysSerPheProLeuThrGluLysGlyPheValAl 113
|||||
1294 TCMACTGCACTTTCATCAATCAATCCGTCATGAGACCGAATAATGTCATC 1343

114 ProAspLeuSerGluHisLysLysAsnProGlyPhe 125
|||||
1344 CCATCTATTGGCGCTTCACATGCTATGCTCCCTCTTT 1379

seq_name: /cgn2_6/ptodata/1/lna/6B.COMB.seq:US-09-371-913A-10

seq_documentation_block:
Sequence 10, Application US/09371913A
Patent No. 6297369
GENERAL INFORMATION:
APPLICANT: Schaeff, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Flockstad Lee, Stacey
APPLICANT: Walz, Mikki
APPLICANT: Sturges, Blake
TITLE OF INVENTION: Pesticidal Toxins and Genes from Bacillus laterosporus
FILE OF INVENTION: Strains
FILE REFERENCE: MA-719XC2 US
CURRENT APPLICATION NUMBER: US/09/371,913A
CURRENT FILING DATE: 1999-08-10
PRIOR APPLICATION NUMBER: 60/095,955
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/138,251
PRIOR FILING DATE: 1999-06-08
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 1041
TYPE: DNA
ORGANISM: Bacillus laterosporus
US-09-371-913A-10

alignment_scores:
Quality: 66.50 Length: 176
Ratio: 0.899 Gaps: 9
Percent Similarity: 42.045 Percent Identity: 21.023

alignment_block:
US-09-728-670-10 x US-09-371-913A-10 ..

Align seg 1/1 to: US-09-371-913A-10 from: 1 to: 1041

```

```

7 GlyLysTyrLysGly.....AspAspAlaSerTyrPheGluPr 20
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8 GGGTATTATTAAAGAAAGATTATGATCTTACCTGTGGTGCAC 57
20 OThrClProTyrLeuMetValAsnValThrGlyValAspSerLysGly 37
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
58 GACA.....CGTGATTA.....CGTGATTA.....CGTGATTA 68
37 snGluLeuLeuSerProHisTyrValGluPheProIleLysProGlyThr 53
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
69 ATACTCTATT.....TATGACCAACAACAGCAAAAT 100
54 ThrLeuThrLysGluLysIleGluTyrTyr.....ValGluTrp.... 66
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 ACACCTAGTAGATCAAAAGCATCAAGAAATCATCTATTCCCTGGATTGG 150
67 .AlaLeuAspAlaThrAlaTyrLysGluPhe..... 76
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 ATTGATTGAGTAGTGCACAGCAGATTTTCACATTTAAATTGCAGATG 200
77 .....ArgValValGluLeuAspProSer..... 84
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 ATGAAATATGCATCATGATGATTTGGATGGAAAGTATTTCGAAAAAGGT 250
85 .....AlaLysIleGlu 88
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 AACCAATAACAAGGTTCATTAGAAAAAGACAGTTGGTCCAAATATAA 300
88 uValThrTyr.....TyrAspLysAsnLysLysGly 99
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 AATTAGATGACCATGACAGATGCAATTACATATGATATATAAACTTTTA 350
99 LuGluThrLysSerPheProIleThrGluLysGlyPheValValProAsp 115
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 AAGACTTAAGTATTCAAGATAGATAGTCAAAATCACTCTCTACAAATT 400
116 LeuSerGluHisIleLysAsnProGlyPheAsn.....Le 127
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 CAACAGATGAACTGAGAAACCCGTGATTATTAAGAAAGAAACGCAAG 450
127 uIleThrLysValValIleGluLysLys 136
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
451 AATTCTTAAGAAAGCATGCAAGCAAA 478

```

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:us-09-313-930-1

seq_documentation_block:

```

; Sequence 1, Application US/09313930
; Patent No. 6235723
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; TITLE OF INVENTION: Antisense Oligonucleotide Modulation of Human Protein
; FILE REFERENCE: ISPH-0357
; CURRENT APPLICATION NUMBER: US/09/313,930
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2104
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59)..(2089)
; PUBLICATION INFORMATION:
; AUTHORS: Artis, J. P.
; AUTHORS: Basta, P. V.
; AUTHORS: Holmes, W. D.
; AUTHORS: Ballas, L. M.
; AUTHORS: Moormaw, C.
; AUTHORS: Rankl, N. B.
; AUTHORS: Blobel, G.

```

```

; AUTHORS: Loomis, C. R.
; AUTHORS: Burns, D. J.
; TITLE: Molecular and biochemical characterization of a
; TITLE: recombinant human PKC-delta family member
; JOURNAL: Biochim. Biophys. Acta
; VOLUME: 1174
; ISSUE: 2
; PAGES: 171-181
; DATE: 1993-08-19
; DATABASE ACCESSION NUMBER: L07860
; DATABASE ENTRY DATE: 1993-11-02
; US-09-313-930-1

alignment_scores:
  Quality: 66.00      Length: 172
  Ratio: 0.917      Gaps: 9
  Percent Similarity: 41.860      Percent Identity: 21.512

alignment_block:
  US-09-728-670-10 x US-09-313-930-1 ..

Align seg 1/1 to: US-09-313-930-1 from: 1 to: 2104

2 SerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTyrPh 18
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86 AACCTATATGAGCTGGCTCCCTGCAGCGCAGGACGAGCGAAGCAG.. 133
18 eGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspSerL 35
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 .....CCCTCTGTCGCGCTG.....A 149
35 ysGluAsnGluLeuLeuSerProHisTyrValGluPheProIleLysPro 51
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
150 AGATGAGAGAGGCGCTCAGC.....ACAGAGCGT 178
52 GlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrp.... 66
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 GGGAAACACTGCTGTCAGAAAGACCGACCATGATCTCTGAGTGAAGTC 228
67 .AlaLeuAspAlaThrAlaTyrLysGluPheArgValVal..... 79
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
229 GACGTTGATGCCACATCTAT...GAGGGCGCGTCATCAGATTGTC 275
79 .....
276 TATATCGCGGACAGAGAGCAGTGTCTGAGTGCAGCGGTGGTGTGCG 325
79 .....
326 GTGCTGCGCCAGCGCTGCAGAGAACATGCAAGGCTGAGTTCTGCT 375
80 .GluLeuAspProSerAlaLysIle.....GluValThrTyrTyr... 92
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
376 GGACTCGAGCTTCAGCGCAGGCTGATGCTGCTCAGTATTTCCCTGG 425
93 .....AspLysAsnLysLysGluGluTyrThrLysSerPhe 104
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
426 AGAGCTGTGATTCGAACAACATCTATGCGAGTGAGAGAGGCGCAATTC 475
105 Pro...IleThrGluLysGlyPheValValProAspLeuSerGluHisI 120
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
476 CCAACGATGAACCGCGCGGACCATCAACAGGCCAAATCCACTACAT 525
120 eLysAsnProGlyPhe 125
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
526 CAAGAACCATGAGTTT 541

```

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:us-09-147-208-26

seq_documentation_block:

```

; Sequence 26, Application US/09147208
; Patent No. 6333303

```

```

GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Antiviral R1c1n-Like Proteins
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERSKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/147,208
FILING DATE: 02-MAR-1999
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Rudolph, John R.
REGISTRATION NUMBER: 38,003
REFERENCE/DOCKET NUMBER: 7841-76
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 9639 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-147-208-26

alignment_scores:
Quality: 66.00 Length: 138
Ratio: 0.930 Gaps: 11
Percent Similarity: 51.449 Percent Identity: 26.812

alignment_block:
US-09-728-670-10 x US-09-147-208-26/rev ..
Align seg 1/1 to reverse of: US-09-147-208-26 from: 1 to: 9639

8 LysTyrLysGlyAspAspAlaSerTyrPheGluProThr..... 21
1438 AAATACAAA.....TCGTTATTGAGCAGACCATCCAAA 1404
22 ....GlyProTyrLeuMetValAsnValThrGlyValAspSerLysGly 37
1403 TGGGGGCCCT.....ACCTCTAACAGCGTCGTGAAAAAGGGCG 1366
37 snGluLeu...LeuSerProHisTyrValGluPheProIleLysProGly 52
1365 ACCGCGTGTATTGACCGCATACAGTG...CTCATTAATAATCCGGC 1319
53 Thr.....ThrLeuThrLysGluLysIleGluTyrTyrVal..... 64
1318 GTGATTCAGCTGATTATGAAAGTCCAAATGCTTACGCCCATAGAAATTACA 1269
65 .GluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValValGluL 81
1268 AGAATGCGCTTTACAGAAGAAATTCCTCAAGTCCTGTGCACCGCGCACT 1219
81 euasProSerAlaLysIleGluValThrTyrTyrAspLysAsnLysLys 97
1218 ACGATCCGCGCATCAAA.....CAACGG 1196
98 LysGluGluThrLysSerPheProIleThrGluLysGlyPheValAlaPr 114
:::|||||:::|||||:::

```

```

1195 GAGGAGAGAGCAAAACAGTTG...CTGACTAAG.....CTGATTGC 1158
114 oasPLeuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysV 131
1157 GACATTTCACCGACAGACAC.....AACGGCGTCGAAGGGG 1123
131 alValIleGluLys 135
1122 TGTGCGCGCAAAA 1109

seq_name: /cgn2_6/plodata/1/lna/6A_COMB.seq:US-08-978-741-5
seq_documentation_block:
Sequence 5, Application US/08978741
Patent No. 6100076
GENERAL INFORMATION:
APPLICANT: Yang Mang, Michael W. Spellman
TITLE OF INVENTION: O-Fucosyltransferase
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,741
FILING DATE: 26-No. 6100076-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/792498
FILING DATE: 31
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1041P1
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 11284 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-978-741-5

alignment_scores:
Quality: 66.00 Length: 138
Ratio: 0.930 Gaps: 11
Percent Similarity: 51.449 Percent Identity: 26.812

alignment_block:
US-09-728-670-10 x US-08-978-741-5/rev ..
Align seg 1/1 to reverse of: US-08-978-741-5 from: 1 to: 11284

8 LysTyrLysGlyAspAspAlaSerTyrPheGluProThr..... 21
1438 AAATACAAA.....TCGTTATTGAGCAGACCATCCAAA 1404
22 ....GlyProTyrLeuMetValAsnValThrGlyValAspSerLysGly 37
1403 TGGGGGCCCT.....ACCTCTAACAGCGTCGTGAAAAAGGGCG 1366
37 snGluLeu...LeuSerProHisTyrValGluPheProIleLysProGly 52
:::|||||:::|||||:::

```

MOLECULE TYPE: CDNA
US-08-852-481-1

```

; CLASSIFICATION: 800
;
; PRIORITY DATA:

```

PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 07/791,921
FILING DATE: 15 NOVEMBER 1991
ATTORNEY/AGENT INFORMATION:
NAME: Siegel, Barbara C.
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1035-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4931
TELEFAX: (302) 773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2675 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: glycine max
STRAIN: Cultivar Wye
CELL TYPE: Colydon
IMMEDIATE SOURCE:
LIBRARY: cDNA to mRNA
CLONE: PC161
US-08-232-079-1

```

```

alignment_scores:
Quality: 65.50      Length: 191
Ratio: 0.851      Gaps: 5
Percent Similarity: 40.314      Percent Identity: 19.372

```

```

alignment_block:
US-09-728-670-10 x US-08-232-079-1 ..

```

```

Align seg 1/1 to: US-08-232-079-1 from: 1 to: 2675

```

```

5  AsplysGlystTyrLysLysGlyAsp.....AspAl 15
   |||:  ::|||:|||||
1436 GACCGTGATCATATACAAAGCTGATTTCCAGAAGATATAGTATGATGC 1485
15  aSetTyrPhe.....GluProThrGlyProTyrL 25
   |||:  ::|||:|||||
1486 TAAATTTTGTGCATCAAGTCTTACAGTGAATGATATCCATAGAGTA 1535
25  euMetValAsnValThrGlyValAspSerLysGlyAsnGluLeuSer 41
   |||:  ::|||:|||||
1536 TTAAGTATAAGCTGTGGCCAGTACCCAAATGCGCAACAAGACTTGAT 1585
42  ProHisTyrValGluPheProIleLysProGlyThrThrLeuThr.... 56
   |||:  ::|||:|||||
1586 GCTGATATCATGAGCAGCAGCAAAACCTGTGCTGCCCTGTTTTCG 1635
56  ..... 56
1636 CTTTTCCTCAGTAAATACAGTGGCAATTGTGGGCTTGCTGAATGA 1685
57  .....LysGluLysIleGluTyrTyrValGluTyrPala 67
   |||:  ::|||:|||||
1686 TTGGTCTCTGTCATTTTAAACAGAGCTTGATATGCGCAGCAAGAC 1735
68  LeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProSe 84
   |||:  ::|||:|||||
1736 TCGAATGCGCTGCTTCTTAAAGTGCACGTGTTAAGATGTCTCTAA 1785
84  rAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGluT 101
   |||:  ::|||:|||||
1786 CAATTGTGTGAGCATTAACCTGGATTAACAATGAGTCAAAACCTGTCA 1835
101  hrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSer 117
   |||:  ::|||:|||||

```

```

1836 CTAAACAGT.....AGGATACACAA 1855
118  GluHisIleLysAsnProGlyPheAsnLeuIle..... 128
   |||:  ::|||:|||||
1856 GAGTAAATGTTGAGACCTGGCCTGAATTAATCAAAATTTCAAGATA 1905
129  .....T 129
1906 TACCAAAAGACATCATCTGATGATTTGGCTTATGAGGCCGTC 1955
129  hrLysValValIleGluLysLys 136
   |||:  ::|||:|||||
1956 AGAAGACTATTTTGGAGAGAA 1978

```

4

5

OM of: US-09-728-670-10 to: EST: * out-format : pfs

Date: Sep 1, 2002 3:56 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frimex+pn.model -DEV=xlh
-Q/cg2.1/USPTO.spool/US09728670/runat_30082002_084912_5292/app-query.fasta.1.196
-DB=EST -OFMT=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -TGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=50 -MODE=LOCAL -OUTFMT=pfs -NORM-ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09728670.ecgsl.1.2488
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPUX -WAIT -THREADS=1

Search information block:

Query: US-09-728-670-10

Query length: 136

Database: EST: *

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 1752.300000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est2:BI207534	+	85.00	163.33	7.51	655 BI207534 EST525574 CTOS Lycoper
gb_est1:AM485868	-	83.00	163.41	7.43	393 AM485868 68470 MARC LPIS Sus sc
gb_gss:AZ052631	+	82.00	159.39	12.45	500 AZ052631 RPT-23.431D1E.TJ RPT
gb_gss:BH248657	+	82.00	156.10	18.87	747 BH248657 AUIUA33F AUIJ Arabid
gb_est2:BI203214	+	81.50	154.59	23.04	793 BG584867 EST486628 MHAM Medicag
gb_est2:BI203214	+	80.00	153.50	26.03	622 BI203214 EST521254 CTOS Lycoper
gb_est2:BI207923	+	80.00	148.83	48.20	1100 BI207923 603060063F1 NIH.MGC.1
gb_est2:BG547905	+	79.00	151.71	33.32	602 BG547905 IPL.6.G12.B1.A002 Immu
gb_est2:BG598438	+	78.50	150.92	36.86	585 BG598438 EST430161 GYSN Medicag
gb_est2:BI205150	+	78.00	151.50	34.26	481 BI205150 HVSMM00111F4F Hordeum
gb_est2:BG581416	+	78.00	148.33	51.39	708 BG581416 EST483149 GVN Medicag
gb_est2:BE455199	+	78.00	148.05	51.62	711 BE455199 HVSME0096304F Hordeum
gb_est2:BF260862	+	78.00	148.05	53.30	733 BF260862 HVSME0023006F Hordeum
gb_est2:BG583222	+	78.00	148.04	53.37	734 BG583222 EST484972 GVN Medicag
gb_gss:CN50723D	+	78.00	145.46	74.29	1006 A1426791 clone BA0AB015B05 of
gb_gss:CN50738U	+	78.00	145.27	76.15	1030 A1427212 clone BA0AB018D05 of
gb_gss:BM483752	+	77.50	146.16	67.94	815 BM483752 BOGW467R BOGW Brassic
gb_est2:BE517562	+	76.50	147.92	54.17	511 BE517562 WHE0628_F09_L18ZA Whea
gb_gss:AO501650	+	76.00	147.99	53.69	447 AO501650 V17E9 mfn-3XHA/lacz In
gb_est1:AA4840229	+	76.00	147.88	54.45	453 AA4840229 VW899406.r1 Stralagene
gb_gss:AZ672674	+	76.00	142.41	110.93	885 AZ672674 ENTU0505R Entamoeba h
gb_gss:BH158255	+	76.00	142.36	110.93	890 BH158255 ENTU0505R Entamoeba h
gb_est2:BE336273	+	75.50	141.98	116.05	822 BE336273 G01062663F1 NIH.MGC.10
gb_est1:AM036063	+	75.00	145.36	75.28	480 AM036063 EST278077 tomato seed,
gb_est1:AM052970	+	75.00	144.20	87.33	553 AM052970 A0052970 Dictyostelium
gb_est2:BM474008	+	75.00	138.71	176.60	1082 BM474008 AGECOURT.6489861 NIH
gb_est1:AA549478	+	74.50	146.29	66.83	378 AA549478 vK6612.s1 Knowles Sol
gb_est2:BI204379	+	74.50	143.65	93.77	522 BI204379 EST522419 CTOS Lycoper
gb_est2:BI206210	+	74.50	143.65	93.77	522 BI206210 EST522975 CTOS Lycoper
gb_est2:BI206210	+	74.50	143.65	93.77	522 BI206210 EST524250 CTOS Lycoper
gb_est2:BI206708	+	74.50	143.65	93.77	522 BI206708 EST524748 CTOS Lycoper
gb_est2:BI207927	+	74.50	143.65	93.77	522 BI207927 EST525967 CTOS Lycoper
gb_est2:BI211112	+	74.50	143.65	93.77	522 BI211112 EST529152 CTOS Lycoper
gb_est2:BI205134	+	74.50	141.50	123.55	679 BI205134 HVSMM0021119F1 Hordeum
gb_est2:BF570018	+	74.50	138.39	184.09	993 BF570018 G02186015T1 NIH.MGC.45
gb_est1:AM036324	+	74.00	143.26	98.88	483 AM036324 EST278449 tomato seed,
gb_est1:AM036351	+	74.00	143.17	99.66	488 AM036351 EST278476 tomato seed,
gb_est2:BI207927	+	74.00	141.31	126.59	613 BI207927 pgrln.ph004.119 Normal
gb_est2:BI205721	+	74.00	141.31	135.49	654 BI205721 pgrln.ph005.k14 normal
gb_est2:BI209102	+	74.00	140.01	149.43	718 BI209102 EST548991 tomato flowe

gb_gss:AZ671442 - 73.50 137.80 198.44 830 | AZ671442 ENTU127R Entamoeba
gb_est1:AL164607 + 73.00 143.72 92.86 355 | AL164607 A065P71U Hydril asp
gb_est2:TI9018 + 73.00 142.71 105.79 402 | TI9018 FI1010T Testis 1 Homo
gb_est2:BF158099 + 73.00 141.66 121.03 457 | BF158099 FI28E10.Y1 Sugano K
gb_est1:AL389514 + 73.00 140.49 140.55 527 | AL389514 MEB55D02R1 MTRC Me

seq_name: gb_est2:BI207534

seq_documentation_block:

LOCUS BI207534 655 bp mRNA linear EST 11-JUL-2001
DEFINITION EST525574 CTOS Lycopersicon esculentum cDNA clone CTOS14114 5' end,
mRNA sequence.
ACCESSION BI207534
VERSION BI207534.1 GI:14685258
KEYWORDS EST.
SOURCE Lycopersicon
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 655)

REFERENCE 1 (bases 1 to 655)

AUTHORS van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R.,
Romling, C. and Tanksley, S.
TITLE Generation of ESTs from Tomato Suspension Cultures
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.

FEATURES

source

1..655

/organism="Lycopersicon esculentum"

/cultivar="TA436, E6203"

/db_xref="taxon:4081"

/clone="CTOS14114"

/clone_lib="CTOS"

/issue_type="suspension cultures"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2:
XhoI; Suspension cultures of L. esculentum E6203 were grown
in Murashige and Skoog based medium, supplemented with 15%
coconut milk (filter sterilized and added after
autoclaving), 2% saccharose, and 1mg/ml 2,4d (pH5.8).
Fresh medium was added every 7 days, and cultures were
grown at 25 C, with 12hrs of light and continuous
shaking."

BASE COUNT 200 a 147 c 132 g 176 t

ORIGIN

alignment_scores:

Quality: 85.00 Length: 136
Ratio: 1.232 Gaps: 6
Percent Similarity: 50.735 Percent Identity: 26.471

alignment_block:

US-09-728-670-10 x BI207534

Align seg 1/1 to: BI207534 from: 1 to: 655

20 ProthrglyProTyrLeuMetValAsnValThrGlyValAspSerLysG1 36
|||||
|||||

270 CCAGACGGGGAATTTGCCCGCTCAAGACACACCCATGAT..... 311
|||||

36 yAsnGluLeuLeuSerProHisTyrValGluPheProIleLysProGlyT 53
:::|||||

312TTCTGAACCCATACGTA.....CGAA 336
:::|||||

53 hrrHrleThrGlyGluLysIleGluTyrTyrValGluTyrPalaLeuasp 69
:::|||||

337 CGAGGATCAAAAACCTCCAAATGATATGACATACATGACACTGCAC 386
:::|||||

70 AlarhAla.....TyrlysgluPh 76
 : : : : :
 387 AGCAGTGCATAAATGAACCCGTGGGAGAGTTTATGATAGAAAGTCCG 436
 76 eayvalValGluLeuasp.....ProserAlaValIleGluValT 90
 : : : : :
 437 AAGAGTGTATGATTAAGAACATTCACCTGGTTCATTCCTACACTT 486
 90 hTyrTyraPlyAsnLysLysLysGluGluThryLysSerPhePro... 105
 : : : : :
 487 CCAATTGGCATTAAGTAAAGAAAGCGTATTTATGATTCGCCCT 536
 106IleThrGluLysGlyPheValValProAspLeuSerGluH 119
 : : : : :
 537 CACGACGATTCGCTTGGAAACTGATGATTCACAGACCTGTGAATCA 586
 119 s.....IleLysAsnProGlyPheAsnLeuIleThrL 130
 : : : : :
 587 CCTTAATTTCCATCTCATTTCTGAATCCGGAGAGATACGTCACAT 636
 130 ysValVal 132
 : : : : :
 637 CAGTGTG 644

seq_name: gb_est1:AW485868

seq_documentation_block:

LOCUS AW485868 393 bp mRNA linear EST 09-JUL-2000
 DEFINITION 68470 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION AW485868
 VERSION AW485868.1 GI:7055974

KEYWORDS EST.

SOURCE pig.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 393)
 Fahnensterg, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.
 and Keele, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine

JOURNAL Unpublished (2000)

COMMENT

Contact: Smith TP
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smithemall.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -m1nscore 20
 and -m1nmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCCTCAGCAGC

Plate: 28 row: N column: 23

Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers

FEATURES

source

1..393

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC 1Pig"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPOR6; Site.1: XbaI; Site.2: XhoI;
 library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."

BASE COUNT 105 a 83 c 120 g 85 t

alignment_scores:

Quality: 83.00 Length: 116
 Ratio: 1.456 Gaps: 5
 Percent Similarity: 49.138 Percent Identity: 25.000

alignment_block:

us-09-728-670-10 x AW485868/rev ..

Align seg 1/1 to reverse of: AW485868 from: 1 to: 393

20 ProthrnglyProtyrLeuMetValAsnValThrGlyValAspSerlysgl 36
 : : : : :
 333 CCAACCGAGCGCATTTTCTAGATGAGGGGCTACTCGACAAACCTGCGG 284
 36 yAsnGluLeu.....LeuSerProHisTyr 45
 : : : : :
 283 TTCACCTTCCAGGAGAGATGCCAAAGCTGAGACTACACCATCTCTG 234
 45 aGlu.....Phe 47
 : : : : :
 233 TCAGAAAGAACTCCGTGTTTCACTCCAGGCTCCCTGAAGAAAGTCTTT 184
 48 ProIleLysProGly.....ThrThrLeuThrLysGlyLysIleGluTy 62
 : : : : :
 183 CCTGTCCAGCCTGCGCATCTCTGCTGTACAGACAG..... 145
 62 rTyValGluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgVal 79
 : : : : :
 144GATCTCGCTGCCACAGACATGACGACGAGGCTC 114
 79 aGluLeuAspProserAlaLysIleGluValThrTyTyrAspLysn 95
 : : : : :
 113 TTTCCTTGATCTTCAGCTCTGTGTGAGCTTCTCTTGTGAGTCTTC 64
 96 LysLysLysGluGlu.....ThrLysSerPheProIleThrGlyLys 109
 : : : : :
 63 GGTACGCGGAGACTCCATCCATCAATCATTCCTCTCTGCAAGG 16

seq_name: gb_gss:A2052631

seq_documentation_block:

LOCUS A2052631 500 bp DNA linear GSS 30-MAR-2000

DEFINITION RPCI-23-413D16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-413D16

ACCESSION A2052631

VERSION A2052631.1 GI:7343790

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 500)

Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akiret,

B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other_GSSs: RPCI-23-413D16.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pliet de Jong

(piet@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html

Plate: 413 row: D column: 16

Seq primer: SP6

Class: BAC ends.

114 OASP 115
|||
327 TGAT 330

Seq_name: gb_est2:BG584867

seq_documentation_block:

LOCUS	BG584867	793 bp	mRNA	linear	EST 11-APR-2001
DEFINITION	EST486628	MHM Medicago truncatula/gnomus versiforme mixed EST library cDNA clone pMHM-19L14 5' end, mRNA sequence.			
ACCESSION	BC584867				

ACCESSION	BG584867
VERSION	BG584867.1
	GI:13599931

SOURCE Medicago truncatula/GIOMUS versiforme mixed EST library.

Eukaryota; mixed EST libraries.

AUTHORS Harrison, M.J., Liu, J., Town, C.D., Van Aken, S., Utterback, T., Cho, J.

TITLE ESTs from roots of Medicago truncatula after colonization with
Clostridium sporosphaera 2001

COMMENT
JOURNAL
Unpublished (2001)
Contact: Harrison M. J.

COMMENT Contact: Harrison M.J.

FEATURES

Source

1. 1/93

```
/organism="Medicago truncatula/Glommus versiforme mixed EST
library"
```

```
/cultivar="Medicago truncatula genotype Al/  
/db_xref="taxon:119092"  
/accession="F57941.10114"
```

```

/clone==pmham-19L14"
/clone_1b="MHAM"
@tissue_type="roots colonized with clones word format"

```

```

/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The library was
/issue_type=Roots colonized with Glomus versiforme
/issue_type=Roots colonized with Glomus versiforme

```

made from a mixture of RNA from each of these stages. /lab host="E. coli strain X101R"

/note="Vector: pBluescript SK-; site_1: EcoRI; site_2: XhoI: cDNA was prepared from polyA+ enriched RNA from

roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with *Glomus versiforme*. The cDNA was

directionally ligated into the Unizap XR vector from Stratagene and packaged using GigaPack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist

helper phage and propagated in XL0LR cells."

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

61.50	Length:	137
0.994	Gaps:	9
53.328	Percent Identity:	34.304

PERCENT IDENTITY: 44.204

.. BG584867

BC584867 from: 1 to: 793

eAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
 ::::|::|::|::|::|::|::|::|::|::|::|::|
 ::::|::|::|::|::|::|::|::|::|::|::|::|

AGAAGCATTGAAACAAAAGGCTTATGGAATATCAGTTA 209

cocoanut milk (filter sterilized and added after autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8). Fresh medium was added every 7 days, and cultures were grown at 25 C, with 12hrs of light and continuous shaking."

BASE COUNT 192 a 140 c 125 g 165 t
ORIGIN

alignment_scores:
Quality: 80.00 Length: 128
Ratio: 1.250 Gaps: 6
Percent Similarity: 50.000 Percent Identity: 26.562

alignment_block:
us-09-728-670-10 x BI203214 ..

Align seg 1/1 to: BI203214 from: 1 to: 622

```

20 ProthrglyProtyrLeumetValasnValthrGlyValaspSerLysG1 36
   |||||
270 CCGACAGGGGAAATGGCCCGTCAGACACACCCCTATGAT..... 311
   |||||
36 YAsnGluLeuSerProHisTyrValGluPheProIleLysProGlyT 53
   |||||
312 .....TTCCTGAACCCCATACAGTA.....GGAA 336
   |||||
53 hrThrLeuThrLysGluLysIleGluTyrTyrValGluThrAlaLeuAsp 69
   |||||
337 GCAGATCAAAAACCTCCAAATGATGATGATCAATCACTACCTGAC 386
   |||||
70 AlaThrAla.....TyrLysGluPhe 76
   |||||
387 AGCAGTCGAAATAAATGAACCCGTCGGAGATTATGATAGAAGTCGG 436
   |||||
76 earGValAlaGluLeuasp.....ProSerAlaLysIleGluValT 90
   |||||
437 AAGAGTCGATGAGTGAAGATGATGACACCTGCTCAATCTACACTT 486
   |||||
90 hrTyrTyrAspLysAsnLysLysLysGluTThrLysSerPhePro... 105
   |||||
487 CCAATGGCATAAAGTAAAGGAAAGGAAAGTGCATTTATGATCCGCC 536
   |||||
106 .....IleThrGluLysGlyPheValAlaProAspLeuSerGluH1 119
   |||||
537 CATGCAGCATGCTTGGAAACTGATGATTCACAGCGCTGGAATCA 586
   |||||
119 s.....IleLysAsnProGly 124
   |||||
587 CCTAATTTTCCATCCTCAATTTGTAATCCGGA 620

```

seq_name: gb_est2:BI769234

seq_documentation_block:

LOCUS BI769234 1100 bp mRNA linear EST 25-SEP-2001
DEFINITION 603060063F1 NIH_MGC_122 Homo sapiens cDNA IMAGE:5209527 5',
mRNA sequence.

ACCESSION BI769234
VERSION BI769234.1 GI:15760812

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (Bases 1 to 1100)

TITLE NIH-MGC http://mgs.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
<http://image.liml.gov>
Plate: LIML1525 row: m column: 16
High quality sequence stop: 739.

FEATURES

source
1..1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5209527"
/clone_11b="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: PCMV-SPORT6;
Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."

BASE COUNT 266 a 329 c 325 g 180 t

alignment_scores:
Quality: 80.00 Length: 123
Ratio: 1.333 Gaps: 5
Percent Similarity: 48.780 Percent Identity: 26.829

alignment_block:
us-09-728-670-10 x BI769234 ..

Align seg 1/1 to: BI769234 from: 1 to: 1100

```

38 GluLeuLeuSerProHisTyrValGluPheProIleLysProGlyThrH 54
   |||||
581 GAACCTTGAGCCCTAAC.....CCAGATGGCAGAGGAGCAGAC 618
   |||||
54 rLeuThrLysGlu.....LysIleGluT 62
   |||||
619 GATCGTCGCGGAGTTGAGCAGCTGATCACTCTTAAGAGCAGCAGACT 668
   |||||
62 yrTyrValGluTrp.....AlaLeuAspAlaThrAlaTyrLys 74
   |||||
669 ATGGGCTCTTGCCCGCCCTTGAGGAGCTTAGACTTGCCATCTCAAT 718
   |||||
75 Glu.....PheArgValAlaGluLeuaspProSe 84
   |||||
719 AGCAATCATGATGTCATCACCATTCCTCTTGCAAACTCACAACCTC 768
   |||||
84 rAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGluT 101
   |||||
769 AGCAAGCTTGACCGCTCACCTTGAAGAGCCAGAGGAGGAGCAGCA 818
   |||||
101 hrLysSerPheProIleThrGluLysGlyPheValAlaProasp..... 115
   |||||
819 CCAAGGAGCTCTTGACAGCACAATTGGGACACACTGTAGCAGGCGCT 868
   |||||
116 .....LeuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrL 130
   |||||
869 GAAGCACTCCGGGATCCGGAAGAGCCCTTGATTAACACGCTTCCAGA 918
   |||||
130 sValValIleGluLysLys 136
   |||||
919 AATCGTCAGAGAGAAAAA 937

```

seq_name: gb_est2:BG947905

seq_documentation_block:

LOCUS BG947905 602 bp mRNA linear EST 12-JUN-2001

Align seg 1/1 to: BE998438 from: 1 to: 585

```

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17
|||||.....:|||||
45 TCAGTGTGCTGTAAGACATTGAAACAAAGGCTTATGAAATATCAGTTA 94
17 PheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||.....:|||||
95 TTTT...CCTGAGAGACC.....ATTACACAAATAGCAT 126
34 eTysGlyAsnGluLeu.....:|||||
127 CAGTACAAATGATGTGTGTGATGAACATGATGTTGAAATGGTTG 176
41 .....SerProHisTyrValGluPheProIleLysPr 51
177 AAAAAACAACACAAATTCGTTTGTATGTTCTTT.....GGAAG 220
51 oGlyThrThrLeuThrLysGlyLysIleGluTyrTyrValGluTyrPal 68
|||||.....:|||||
221 TGTGTGAACACTTCTCAAGACAAATCAAT...GAGATAGCTTTGGTT 267
68 euAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProSer 84
|||||.....:|||||
268 TGGAAATGAGTGTGCAAAAGATTCATTGGGTTGTGAGGCA...CCAAGT 314
85 AlAlaLysIleGluValThrTyrTyrAspLysAsnLysLys..... 97
315 GATTCAAGTCAATGACACCTTATCTTGAATCTGCANATGAAATGATCCTTGA 364
98 .....LysGluGluThrLysSerPheProIleThrG 108
365 ATTCTTCGACGAAGGTTTCACGAAGACCAAA.....G 399
108 LuLysGlyPheValValPro.....AspLeuSerGlu 118
|||||.....:|||||
400 AAAAAAGTTTCATTTTGCATCATGCGCACCTCAAGTTGAATACTTAA 449
119 HisLeuLysAsnProGlyLys 125
|||||
450 CATAGTTCAGTTGTGTGATTT 470

```

seq_name: gb_est2:B1953150

seq_documentation_block:

LOCUS B1953150 481 bp mRNA linear EST 19-OCT-2001

DEFINITION HVSMEM0011F14f Hordeum vulgare green seedling EST library

HVCNDA0014 (Blumeria infected) Hordeum vulgare cDNA clone

HVSMEM0011F14f, mRNA sequence.

ACCESSION B1953150

VERSION B1953150.1 GI:16297584

KEYWORDS EST.

SOURCE

ORGANISM

Barley.

Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

; Triticeae; Hordeum.

1 (bases 1 to 481)

Wing,R., Close,T.J., Kleinof,A., Wise,R., Chin,A., Begum,D.,

Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons

J., Oates,R. and Main,D.

Development of a genetically and physically anchored EST resource

for barley genomics: Blumeria infected Morex (compatible) seedling

cDNA library

Unpublished (2001)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: twing@clemson.edu

Total hg bases = 436
Seq primer: AATTACCTCCTCAAGG
High quality sequence stop: 475.
Location/Qualifiers

FEATURES

source

1.481

/organism="Hordeum vulgare"

/cultivar="Morex"

/db_xref="taxon:4513"

/clone="HVSMEM0011F14f"

/clone_lib="Hordeum vulgare green seedling EST library

HVCNDA0014 (Blumeria infected)"

/tissue_type="green seedling leaf"

/lab_host="TUC121"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:

XhoI; Morex (mla) plants were greenhouse grown in the R

Wise lab at Iowa State University, Ames, IA; 7 day old

green seedlings were infected with isolate 5874 of

Blumeria graminis f. sp. hordei, and leaves were harvested

24, 48 and 72 hr post-inoculation and snap frozen (Wise).

In the TUC close lab at the University of California,

Riverside, total RNA was prepared from each sample pool,

equal quantities of all three RNA pools were combined,

poly(A) RNA was purified from the mixture, one primary

unamplified cDNA library was made, and 1 million p1u were

in vivo excised to give pBluescript SK(-) cDNA phagemids

(Chin). Phagemids were plated and picked at the Clemson

University Genomics Institute (UCGI) (Begum, Palmer,

Frisch, Atkins and Wing). Plasmid DNA preparations, DNA

sequencing and sequence analysis were performed at UCGI

(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).

The sequence has been trimmed to remove vector sequence

and contains a minimum of 100 bases of phred value 20 or

above. For more details on library preparation and

sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order

this clone see http://www.genome.clemson.edu/orders Also

see Close et al, Wing R, Kleinof A, Wise R (2001)

Genetically and physically anchored EST resources for

barley genomics. Barley Genetics Newsletter 31:29-30.

(http://wheat.pw.usda.gov/ggpages/bgrn/31/cover.html)"

BASE COUNT 136 a 81 c 121 g 142 t 1 others

ORIGIN

alignment_scores:

Quality: 78.00 Length: 116

Ratio: 1.393 Gaps: 5

Percent Similarity: 48.276 Percent Identity: 25.862

alignment_block:

US-09-728-670-10 x B1953150 ..

Align seg 1/1 to: B1953150 from: 1 to: 481

```

16 SerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyVa 32
:::|||||
74 GCCATATTTCGAAGATG...CCGTGTTGCGACTTCTTCTCT..... 115
32 LAspSerLysGlyAsnGluLeuSer..... 41
|||||.....:|||||
116 GACTCTGCAAGAGCTTAAACCTTGATGACGCGTTGAGGTCAACGTA 164
42 ..ProHisTyrValGluPheProIleLysProGlyThrThrLeuThrLys 57
|||||.....:|||||
165 TTCACACCTTGTCTTCTTCAATGCAAAAGCTGTGAAGTTCTTACGAT 214
58 GluLysIleGluTyrTyrValGluTyrPalalaLeuAspAlaThrAlaTyr 74
|||||.....:|||||
215 GAAGGAGTTGAGTTGTAAGATGATGATGATGATGATGATGATGATGAT 264
74 sGluPheArgValValGluLeuAspProSerAlaLysIleGluValThrT 91
|||||.....:|||||
265 AACTGAGAGATCAATGATGATGATGATGATGATGATGATGATGATGAT 295

```

```

91 yrrttrapsblysanlyslgslgluThr..... 101
    ::||| ::||| ::|||
296 .....GAAAAGGACGCTAACGATATCAACTATTCATGCTGCTGGT 340
102 .....LysSerPheProIleThrGluLysGlyPheValAlaPro 114
    ::||| ::||| ::|||
341 ACAGCAAAATGCTGCTACGCTAATTCACACGCGGAGAGAGAGTACC 388

seq_name: gb_est2:BG581416
seq_documentation_block:
LOCUS BG581416 708 bp mRNA linear EST_11-APR-2001
DEFINITION EST1683149 GVN Medicago truncatula cDNA clone pGVN-64F2 5' end, mRNA
sequence.
ACCESSION BG581416
VERSION BG581416.1 GI:13596480
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosida I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE 1 (bases 1 to 708)
AUTHORS Fedorova,M., Pletson,B.L., Samac,D.A., Vance,C.P., Gentl,G.S., Town
C,D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago
truncatula, 2001
JOURNAL Unpublished (2001)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
University of Minnesota name: M382475e TIGR sequence name:
MTCU225TK More information is available at: http://www.medicago.org
Seq primer: Skmod (CTA gAA CTA gTg gAT CC).
Location/Qualifiers
FEATURES
source 1..708
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-64F2"
/clone_1lb="GVN"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month
post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XL0R"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
effective root nodules harvested one month post
inoculation with Sinorhizobium meliloti. The cDNA was
directionally ligated into the Uni-ZAP XR vector from
stratagene and packaged using GigaPack III cold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-ZAP phage using Ex-Absist
helper phage and propagated in XL0R cells."
BASE COUNT 209 a 150 c 179 g 170 t
ORIGIN

```

```

alignment_scores:
Quality: 78.00 Length: 113
Ratio: 1.393 Gaps: 1
Percent Similarity: 49.558 Percent Identity: 23.894

```

```

alignment_block:
US-09-728-670-10 x BG581416 ..
Align seg 1/1 to: BG581416 from: 1 to: 708

```

```

12 gIyaspapAlaSerTyrrPheGluProThrGlyProTyrlleuMetValas 28
    ||||| ::||| ::|||
345 GGTGATGATGACGCTGTTTGTCTGATGCTCGCATTCCTT..... 386
28 nValthrGlyValaLysSerLysGlyAsnGluLeuLeuSerProHisTyrv 45
    ::||| ::|||
387 .....CCTCCACCTG 396
45 aIGluPheProIleLysProGlyThrThrLeuThrLysGluLysIleGlu 61
    ||||| ::||| ::|||
397 GTGAGATGAGACGCGGAGAGAGTACGCTTCGCGGATGGCGCGCTGA 446
62 TyrrTyrrValGluTrpAlaLeuAspAlaThrAlaTyrrLysGluPheArg 78
    ::||| ::|||
447 AATGCCATTGACCTAGACGAGAGAGAGAAAGCGAAAGAAATGAGATT 496
78 lValGluLeuAspProSerAlaLysIleGluValThrTyrrAspLysA 95
    ::||| ::|||
497 GAAGATTATTGAGAGAGGCTGAGAGAAATATAGGTGCTTCTATGAGAAA 546
95 snLysLysLysGluLysThrLysSerPheProIleThrGluLysGlyPhe 111
    ::||| ::|||
547 GGAAGCTTAATGTGACACTACAGACGTTCAAAATAGAGAAAGGCGAGAA 596
112 ValValProAspLeuSerGluHisIleLysAsnProGly 124
    ||||| ::|||
597 GTTGTTCGTAATCAAGACGACTTACACAAAGAGGC 635

seq_name: gb_est2:BE455199
seq_documentation_block:
LOCUS BE455199 711 bp mRNA linear EST_22-OCT-2001
DEFINITION HVSMEN0096J04f Hordeum vulgare 5-45 DAP spike EST library
HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEN0096J04f,
mRNA sequence.
ACCESSION BE455199
VERSION BE455199.2 GI:13189831
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 711)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Fritsch,D., Yu
Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
R.D., Close,S.J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics: Morex 5-45 DAP spike cDNA library
JOURNAL Unpublished (2001)
COMMENT On Jul 26, 2000 this sequence version replaced g1:9464636.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total hg bases = 568
Seq primer: AATTAACCTCACTAAAGGG
High quality sequence stop: 704.
Location/Qualifiers
FEATURES
source 1..711
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEN0096J04f"
/clone_1lb="Hordeum vulgare 5-45 DAP spike EST library
HVCNDA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/note="Vector: LambdaZAP; Site_1: EcoRI; Site_2: XhoI;

```

Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 5, 10, 15, 20, 30 and 45 DAP (Fenton). Total RNA was prepared from each pool, equal quantities of all six RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give plasmids SK(-) cDNA phagmids (Choi) in the TJ Close lab at the University of California, Riverside. Phagmids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations at CUGI sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

<http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinohs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html>)

BASE COUNT 194 a 134 c 186 g 197 t

ORIGIN

alignment_scores:

Quality: 78.00 Length: 116
Ratio: 1.393 Gaps: 5
Percent Similarity: 48.276 Percent Identity: 25.862

alignment_block:

US-09-728-670-10 x BE455199 ..

Align seg 1/1 to: BE455199 from: 1 to: 711

```

16 SerTyrPheGluProThrGlyProTyrLeuMetValaenValThrGlyVa 32
   :::::||||| :::::|||||
211 GCCCTATTTTGCMAAGATG...CCGTGGTTGGCAGTTCTTCTCT... 252
32 LasperlysglyanngluLeuSer..... 41
   |||||:::|||||:::
253 .GACTCTGAAGACGTAAACCTGATGAGGGTTGAGGTCAACGGTA 301
   |||||:::|||||:::
42 ..ProHisTyrValGluPheProIleLysProGlyThrThreunThyls 57
   |||||:::|||||:::
302 TTCACACCTTCTTTCCTGTCGCAAAACTGGGAAGTCTTACTGAT 351
   |||||:::|||||:::
58 GluTysIleGluTyrValGluTrrPalaleuAspAlarhAlatryLy 74
   |||||:::|||||:::
352 GAAAGAGTTGAGTTGTGAAGTATATGTAAGAACTTATCCTTTAC 401
   |||||:::|||||:::
74 sGluhearyValValGluLeuAspProserAlalysIleGluValThrT 91
   |||||:::|||||:::
402 AACTGAGAGGATCAATGAATTAAGAACAA..... 432
   |||||:::|||||:::
91 yTyrAspLysanlyslslysgluGluThr..... 101
   |||||:::|||||:::
433 .....GAAAAGCAGCTAAGATATCAAACTATTCATAGTCTTGT 477
   |||||:::|||||:::
102 .....LysSerPheProIleThrGluLysGlyPheValaPro 114
   |||||:::|||||:::
478 ACAGCAAAATCGTCTAGTATTTCAAAACGAGGAGAGGATGCC 525
   |||||:::|||||:::

```

seq_name: gb_est2:BF260862

seq_documentation_block:

LOCUS BF260862 733 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEF0033C06f Hordeum vulgare seedling root EST library HVCNMA0007
(Etiolated and unstressed) Hordeum vulgare cDNA clone
HVSMEF0033C06f, mRNA sequence.

ACCESSION BF260862
VERSION BF260862.2 GI:13121015
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare

REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Hordeum.
1 (bases 1 to 733)
Wing,R., Close,T.J., Kleinohs,A., Wise,R., Begum,D., Frisch,D., Yu ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton ,R.D., Oates,R. and Main,D.

TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library
JOURNAL Unpublished (2001)
COMMENT On Nov 16, 2000 this sequence version replaced gi:11189975.

CONTACT: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 535
Seq primer: AATTAACTCCTCACTAAGG
High quality sequence stop: 722.

FEATURES

source

Location/Qualifiers

1..733
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEF0033C06f"
/clone_1lb="Hordeum vulgare seedling root EST library HVCNMA0007 (Etiolated and unstressed)."
/tissue_type="Seedling root"
/lab_host="TUC121"
/note="Vector: LambdaZAP; Site_1: EcoRI; Site_2: XhoI; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give plasmids SK(-) cDNA phagmids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagmids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinohs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html>)

BASE COUNT 181 a 196 c 184 g 172 t

ORIGIN

alignment_scores:

Quality: 78.00 Length: 116
Ratio: 1.393 Gaps: 5
Percent Similarity: 48.276 Percent Identity: 25.862

alignment_block:

US-09-728-670-10 x BF260862 ..

Align seg 1/1 to: BF260862 from: 1 to: 733

```

16 SerTyphgclupProthrglyProTyrlLeuMetValAsnValhrglylva 32
   |||||
396 GCGTATTTTCCAAAGATG...CCGTGGTGGAGATTCTTCTCT... 437
32 laepSerlygLyAsngluleuSer..... 41
438 .GACTGGAAGGAGCTAAACCTTGATGAGCGCTTGAGCTCAACGGA 486
42 .ProhIstYrValGluPherProIlelySerProGlyThrThrLeuThrlys 57
   |||||
487 TTCACACACCTGTTCTTCTGATGCAAAAAGTGTGAGTCTTACTGAT 536
58 GlulysIleGluTyTyValGluTTPalaLeuAspAlaThraLayrly 74
   |||||
537 GAAGAGAGTGTGTTGTAAGTAAATGCTATAGAGATCTTCTTTTAC 586
74 sgluPharValValGluLeuAspProSerAlaLysIleGluValThrt 91
   |||||
587 AACTGAGAGGATCAATGAAATTTGAAGCAAA..... 617
91 YrTyraPlyAsnLyLyValGluGluThr..... 101
618 ....GAAAGGCGAGCTAAGATATCAACTTCTAGTGTGCTGGT 662
102 .....LysSerPheProIleThrGluLyGlyPheValPro 114
663 ACAAGCAATGCTGCTACGTAATTTTCAACACGCGAAGAGGTACC 710
seq_name: gb_est2:BG583222

```

seq_documentation_block: 734 bp mRNA linear EST 11-APR-2001
LOCUS BG583222
DEFINITION EST1484972 GVN Medicago truncatula cDNA clone pGVN-7283 5' end, mRNA sequence.
ACCESSION BG583222
VERSION BG583222.1 GI:13598286
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 734)
Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town
C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
ESTs from one month old nitrogen-fixing root nodules of Medicago
truncatula, 2001
JOURNAL Unpublished (2001)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
University of Minnesota name: M3842986 TIGR sequence name:
MTCCE26TK More information is available at: http://www.medicago.org
Seq primer: SKMD (CTA GAA CTC gat CC).
FEATURES
Source Location/Qualifiers
1..734
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-7283"
/clone_1ib="GVN"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month
post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XL0R"
/note="Vector: pBluescript SK-, Site_1: EcoRI; Site_2:

alignment_scores:
Ratio: 78.00 Length: 113
Percent Similarity: 49.558 Percent Identity: 23.894
Gaps: 1
alignement_block:
US-09-728-670-10 x BG583222 ..

Align seg 1/1 to: BG583222 from: 1 to: 734

```

12 GlYAspAspAlaSerTyphgclupProthrglyProTyrlLeuMetValAs 28
   |||||
307 GGTGATGATGACGGTGTCTTCTGATGATGCTCGATCTT..... 348
28 nValThrglyValAsSerlygLyAsngluleuSerProhIstYrV 45
   |||||
349 .....COTCCACCTG 358
45 aGluPherProIlelySerProGlyThrThrLeuThrLyGluLysIleGlu 61
   |||||
359 GTGAGATGAGTCGGAGGAGCTTACGCTTCGCGCAATGGCCGCTCAA 408
62 TyTyTyValGluTTPalaLeuAspAlaThraLayrlyGluPharVal 78
   |||||
409 AATGCCATTGACCTAGAGAGAGCAAGCAAAAGGCAAAAGAAATGATTT 458
78 lValGluLeuAspProSerAlaLysIleGluValhrglyTyraPly 95
   |||||
459 GAAGATTATTGAGGAGCGTGAAGATATAAGTGCTTCTATGAGAAA 508
95 snLyLyLyGluGluThrThrLysSerPheProIleThrGluLysGly 111
   |||||
509 GGAAGCTTAATGTTGACTAACAGGTTCAAAATATGAAAGGCGAGAA 558
112 ValValProAspLeuSerGluHisIleLysAsnProGly 124
   |||||
559 GTTGTTCGATGCTATCAAGAGAACTTCACAAAGAGGC 597
seq_name: gb_gss:CNS072XD

```

seq_documentation_block: 1006 bp DNA linear GSS 07-JUL-2001
LOCUS CNS072XD
DEFINITION clone BA0AB015B05 of library BA0AB from strain CLIB 210 of
KluYveromyces lactis, genomic survey sequence.
ACCESSION AL426791
VERSION AL426791.1 GI:12209985
KEYWORDS GSS.
SOURCE KluYveromyces lactis.
ORGANISM KluYveromyces lactis.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; KluYveromyces.
REFERENCE 1 (bases 1 to 1006)
Bolognin-Fukuhara,M., Toffano-Nloche,C., Artiguenave,F.,
Duchateau-Nguyen,G., Lemaire,M., Marmisele,R., Montrocher,R.,
Robert,C., Ternier,M., Wincker,P. and Wesolowski-Louvel,M.
Genomic exploration of the hemiascomycetous yeasts 11.
JOURNAL FEBS Lett. 487 (1), 66-70 (2000)
MEDLINE 20584721
REFERENCE 2 (bases 1 to 1006)
AUTHORS Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,

XhoI: cDNA was prepared from polyA+ enriched RNA from
effective root nodules harvested one month post
inoculation with Sinorhizobium meliloti. The cDNA was
directionally ligated into the Uni-ZAP XR vector from
stratagene and packaged using GigaPack III Gold packaging
extracts. Plasmids containing GigaPack III Gold packaging
from the recombinant lambda-ZAP phage using Ex-Absist
helper phage and propagated in XL0R cells.."

by Clontech using a combination of random and oligo dT primers. Library was plated and archived by Russell Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 129 a 130 c 114 g 138 t

alignment_scores: Quality: 76.50 Length: 79
Ratio: 1.821 Gaps: 3
Percent Similarity: 53.165 Percent Identity: 30.380

alignment_block:
US-09-728-670-10 x BE517562/rev ..

Align seg 1/1 to reverse of: BE517562 from: 1 to: 511

16 SerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyVal 32
:::|||||
283 GCGTATTGCGAAGATG...CCGTGGTGGCAGTCTCTCTCC... 242
32 AspSerIysGlyAsnGluLeuSer..... 41
241 .GACTGTAAAGCCGTAAAGCCTTGATGACGCGTTGAGCTCAACGTA 193
42 ..ProHisTyrValGluPheProIleLysProHisTyrThrLeuThrLys 57
|||||
192 TTCACACCTGTATCTTCTTGATGCAAAACTGGCAAGTCTTACGAT 143
58 GluIysIleGluTyrTyrValGluThrPalLeuAspAlaThrAlaTyrLys 74
|||||
142 GAGAGAGTGGATTGTGAGTGAATATGATAGACCTTATCTTTAC 93
74 SGIuPheArgValAlaGluLeuAspProSerAlaLys 86
92 AACTGAGAGCAATGATTAAGCAAGCAAGAAAG 56

seg_name: gb_gss:A0501650

seg_documentation_block:

LOCUS A0501650 447 bp DNA linear GSS 29-APR-1999
DEFINITION V17E9 m7n-3kHA/lacZ Insertion library Saccharomyces cerevisiae
genomic 5', DNA sequence.

ACCESSION A0501650
VERSION A0501650.1 GI:4707300

KEYWORDS GSS.
SOURCE Baker's yeast.
ORGANISM Saccharomyces cerevisiae

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 447)

AUTHORS Ros-Macdonald, P., Roemer, T., Coelho, P.S.R., Agrawal, S., Kumar, A.,
desBergues, S.A., Cheung, K.-H., Sheehan, A., Symoniacis, D., Jensen, R.,
Umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
Hager, K., Miller, P., Roeder, G.S., and Snyder, M.

Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
Unpublished (1999)

JOURNAL Contact: Kumar A

COMMENT Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161

FEATURES Email: anuj.kumare@yale.edu
te of m7n-3kHA/lacZ Insertion.
Seq primer: GGCCTTCTTCTTGGCAAGTAC
Class: transposon-tagged.

location/Qualifiers
1..447
/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"
/clone_lib="m7n-3kHA/lacZ Insertion Library"
/lab_host="E. coli"

/note="Vector: pHS56-Sal: A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in pHS56-Sal:
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a m7n-3kHA/lacZ
multitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 138 a 108 c 78 g 122 t 1 others

alignment_scores: Quality: 76.00 Length: 127
Ratio: 1.310 Gaps: 6
Percent Similarity: 45.669 Percent Identity: 22.047

alignment_block:
US-09-728-670-10 x A0501650/rev ..

Align seg 1/1 to reverse of: A0501650 from: 1 to: 447

4 PheAspIysGlyLysTyrLysGlyAspAspAlaSerTyrPheGluPr 20
|||||
284 TTACTTCGCGAAGATTCAAGTCGCGAT..... 255
20 oTHngLpProTyrLeuMetValAsnValThrGlyValAspSerLysGly 36
254CATATTTCGCGTGGTGAATTTTACGCGT 227
37AsnGluLeuLeuSerProHisTyrValGluPheProIle 49
|||||
226 CCGTTTGAAGAAATTCCTAATTTCAAGCATAAGATTTTGGCCACA 177
50 LysProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTr 66
:::|||||
176 AGATCAAGAGGTACAAATCACT.....TG 154
66 PalLeuAspAlaThrAlaTyr.....LysGluPheArgValValG 80
|||||
153 GATTGCTCCAGCTGGTGGTACACTTTGATGAGAACATTTTGAACCTTG 104
80 IuLeuAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnLys 96
|||||
103 AATTGAT...GGCAAGAGTTGATTCCTTTTAC..... 69
97 LysLysGluGluThrLysSerPheProIleThrGluLysGlyPheValVa 113
68CATCTTGCGCTGTGGTTCGCGGCTTGAACCTT 34
113 lProAspLeuSerGluHisIleLysAsnPro 123
33 CAGTGAACGAAACCTCAAGTTAAGGCCCA 3

seg_name: gb_est1:AA840229

seg_documentation_block:

LOCUS AA840229 453 bp mRNA linear EST 27-FEB-1998
DEFINITION vW89a06.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1262098 5' similar to gb:D10214 Mouse mRNA for prolactin
receptor (MOUSE);, mRNA sequence.

ACCESSION AA840229
VERSION AA840229.1 GI:2915888

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 453)
AUTHORS Maria, M., Hillier, L., Allen, M., Bowles, M., Dierlich, N., Dubnue, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

TITLE
 The WashU-HMI Mouse EST Project
 JOURNAL
 Unpublished (1996)
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:664650
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 420.

FEATURES

source
 Location/Qualifiers
 1..453

/organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:1262098"
 /clone_lib="Stratagene mouse skin (#93713)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
 /Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
 dT. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb. Uni-ZAP XR Vector. -5'
 adaptor sequence: 5' GAAATCGCCGACAG 3' -3' adaptor
 sequence: 5' CTCGAGTTTATTTTTTTTTT 3'."

BASE COUNT 110 a 129 c 108 g 106 t
 ORIGIN

alignment_scores:

Quality: 76.00 Length: 120
 Ratio: 1.382 Gaps: 6
 Percent Similarity: 45.833 Percent Identity: 25.000

alignment_block:
 US-09-728-670-10 x AA840229 ..

Align seg 1/1 to: AA840229 from: 1 to: 453

```

18 PheGluProThrGlyProTyr.....LeuMetValAsnValThrGlyVa 32
   ||| |||||::: |||||::: |||||::: |||||::: |||||::: ||
164 TTTCCCCCTTGACCTGTGACCTGTGACCTGTGACCTGTGACCTGTGACCT 213
   ||| |||||::: |||||::: |||||::: |||||::: |||||::: ||
32 IaapSerLygLyAsnGluLeuLeuSerPro...HlaTyRValGluPhep 48
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
214 GATGACAAATGAGGACGACGCTAATGCCATGCCATGCCAATGCCAAGACTATC 263
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
48 ro.....IlelyProGlyThrThrLeuThrlysgLysIleGlu 61
   ||| |||||::: |||||::: |||||::: |||||::: |||||::: ||
264 CGGGTCAGAGGTGTTAAACC..... 283
62 TyTyValGluTrrPalalLeuAspAlaThrAlaTyRlysgLupheargVa 78
283 ..... 283
78 lValGluLeuAspProSerAlaLysIleGluValThrTyTyAspLys. 94
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
284 .ACACACATGATCTGACGACGCTGATGATGACGACGCTGATGACGACGAC 332
95 .....AsnLyLygLyGluGluThrlySerPheProIleThr 107
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
333 ATTCTCTTTTGTCTGAAGATGTGAGAGCCCGAGGCTTAACCC..... 376
108 GlLysGlyPheValValProAspLeuSerGluHisIleLysAsnProGl 124
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
  
```

377 ..CCTGCTTCCACANCCCTGAGATCATGAGAACCGAGAAATCCTGA 423
 124 yPheasnleu 127
 |||||:::
 424 GGCAAAATAT 433

seq_name: gb.gss:A2672674

seq_documentation_block:

LOCUS A2672674 885 bp DNA linear GSS 14-DEC-2000
 DEFINITION ENTLM05TR Entamoeba histolytica sheared DNA Entamoeba histolytica

ACCSSION A2672674
 VERSION A2672674.1 GI:11809820

KEYWORDS
 SOURCE
 ORGANISM

Entamoeba histolytica.
 Entamoeba histolytica.
 Eukaryota; Entamoebidae; Entamoeba.
 1 (bases 1 to 885)

REFERENCES
 AUTHORS
 TITLE
 JOURNAL

CONTACT: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Class: Shotgun
 Seq primer: M13-Reverse
 High quality sequence start: 54
 High quality sequence stop: 758.

FEATURES

source
 Location/Qualifiers

1..885
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica sheared DNA"
 /note="Vector: pROSI; Site_1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaundin and B.
 Batell, Oxford University Press, 1999)."
 Batell, Oxford University Press, 1999).".

BASE COUNT

ORIGIN 410 a 66 c 182 g 227 t

alignment_scores:

Quality: 76.00 Length: 118
 Ratio: 1.134 Gaps: 7
 Percent Similarity: 56.780 Percent Identity: 34.746

alignment_block:
 US-09-728-670-10 x A2672674 ..

Align seg 1/1 to: A2672674 from: 1 to: 885

```

19 GluProThrGlyProTyrLeuMetValAsnValThrGlyValaapSerLy 35
   ||| |||||::: |||||::: |||||::: |||||::: |||||::: ||
94 GAAATGGAAGGAAATTAATCTCAAGTGAAGTGAAGTGAAGGAAACAAA 143
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
  
```



```

35  sGlyAsnGluLeuSerProHisTyrValGluPheProIleLysProG 52
144  A...AAAAATTAATAAAGAAAAAGAAAGAAATGAAATTAAG... 186
52  LyThrThrLeuThrLysGluLysIleGluTyrTyrValGluThrPalau 68
187  .....AAGAAAAA.....GATGTAGAAAAAGCTTAA 213
69  AspaLarThrAlaTyrLysGluPheArgValAlaGluLeuAspProSerAl 85
214  AGAGCAAAAGAAATATATGAAATTCGT.....CATTTATCAAAATGCG 257
85  aLysIleGluValThrTyrTyrAspLysAsnLysLysGluGluThrL 102
258  TAAATTAGAGAGTGC.....CAAAAACAAAAAAATGTAATTTTGT 301
102  ySerPheProIleThrGluLysGluPheValProAspLeuSerGlu 118
302  AAGCATTTCTTAATCTTGAGAG.....ATTAAAGATTAAAGAA 342
119  HisIleLysAsnProGlyPheAsnLeuIleThrLysValAlaIleGlu 135
343  AGGTTTGATGACTTGGTGTATTAAGCAACT...ATTTAATGAGAAA 389
135  sLys 136
390  AAAA 393

seq_name: gb_gss:BH158255

seq_documentation_block:
LOCUS BH158255 890 bp DNA linear GSS 24-SEP-2001
DEFINITION ENT114TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION BH158255
VERSION BH158255.1 GI:15731693
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 890)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HML:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b1loftus@tigr.org
Clones are derived from che Entamoeba histolytica HML:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 10
High quality sequence stop: 559.
Location/Qualifiers
1..890
/organism="Entamoeba histolytica"
/strain="HML:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: PHOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
```

```

H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999).
BASE COUNT 225 a 151 c 92 g 422 t
ORIGIN

alignment_scores:
Quality: 76.00 Length: 118
Ratio: 1.134 Gaps: 7
Percent Similarity: 56.780 Percent Identity: 34.746

alignment_block:
us-09-728-670-10 x BH158255/rev ..

Align seg 1/1 to reverse of: BH158255 from: 1 to: 890

19  GluProThrGlyProTyrIleMetValAsnValThrGlyValAspSerly 35
523  GAAGTTGAGGAAAAAATCTTCAAGTTGAGTGAATCTGAGAGAAACAA 474
35  sGlyAsnGluLeuSerProHisTyrValGluPheProIleLysProG 52
473  A...AAGAAATTAAAAAAGAAAAAGAAAGAAAGAAATTTAAAG... 431
52  LyThrThrLeuThrLysGluLysIleGluTyrTyrValGluThrPalau 68
430  .....AAGAAAAA.....GATGTAGAAAAAGCTTAA 404
69  AspaLarThrAlaTyrLysGluPheArgValAlaGluLeuAspProSerAl 85
403  AGAGCAAAAGAAATATATGAAATTCGT.....CATTTATCAAAATGTCG 360
85  aLysIleGluValThrTyrTyrAspLysAsnLysLysGluGluThrL 102
359  TAAATTAGAGAGTGC.....CAAAAACAAAAAAATGTAATTTTGT 316
102  ySerPheProIleThrGluLysGluPheValProAspLeuSerGlu 118
315  AAGCATTTCTTAATCTTGAGAG.....ATTAAAGATTAAAGAA 275
119  HisIleLysAsnProGlyPheAsnLeuIleThrLysValAlaIleGlu 135
274  AGGTTTGATGACTTGGTGTATTAAGCAACT...ATTTAATGAGAAA 228
135  sLys 136
227  AAAA 224

seq_name: gb_est2:BE536273

seq_documentation_block:
LOCUS BE536273 822 bp mRNA linear EST 09-AUG-2000
DEFINITION M01062663F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3449271 5',
mRNA sequence.
ACCESSION BE536273
VERSION BE536273.1 GI:9764918
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLM8425 row: m column: 16
 High quality sequence stop: 472.

FEATURES

source
 1. 822
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3449271"
 /clone_1lb="NH.MGC_10"
 /cell_line="MGC36"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."
 BASE COUNT 119 a 213 c 287 g 203 t
 ORIGIN

alignment_scores:
 Quality: 75.50 Length: 148
 Ratio: 1.180 Gaps: 5
 Percent Similarity: 43.243 Percent Identity: 22.297

alignment_block:
 US-09-728-670-10 x BE536273/rev ..

Align seg 1/1 to reverse of: BE536273 from: 1 to: 822

```

22 G1PProTyleuMeTValenValThGlyVal..... 32
|||||:|||||:|||||:|||||:|||||:|||||:
643 GCGCCCGACGCTTCTAGCAAAATGAATGGGCTGCCACCAACATTTGGG 594
33 .....AspSerLysGlyAsnGluLeuLeuSerPro.... 42
593 GCGCGTGTGCCACACCAACAGTGTGGGGCCACTCTTAGCCACGACT 544
43 .....HisTyrValGluPheProIleLys 50
543 CAGAGAGCTGGCAAGAGTCGCAACACCTATACCGGGCCAAAGAG 494
51 ProGlyThrThrLeuThrLysGluLysIleGlu..... 61
493 CTGGGCGATAAAAAGCAGAGGAGAACTGAGCTGGAACCTCCCTCC 444
62 .....TyrTyr 64
443 TCCCCACCTCGGGGTGATAGACGAGAAAAATAACCTCCCTCTGGC 394
64 A1G1UTPALaleuAspAlaThrAlaTyrLysGluPheArgValValGlu 80
393 TCCAGAGCGCGCGCGCGCGCGCGCGCGCATCATCTGTTATATATCC 344
81 LeuAspProSerAlaLysIleGluValThrTyrTyrAspLysAlaLys 97
343 CCGGACCGCTCCAGCCCTGAGGCTTGAAGTTAAAAAAA 294
97 alysglglutThrLysSerPheProIleThrGluLysGlyPheValValP 114
293 AAAAGACAAAAGTTACAGTATCTT...ACCCAGACACAGTTTAACTT 247
114 roAspLeuSerGluHisIleLysAsnPro...GlyPheAsnLeu 127
246 CAGAGCTTAAAAAATCTGTCCGCGCGCGCGAGGATGAAATC 203

```

seq_name: gb_est1:AM036063

seq_documentation_block:

LOCUS AM036063 480 bp mRNA linear EST 18-MAY-2001
 DEFINITION EST778077 tomato seed, TAMU Lycopersicon esculentum CDNA clone
 CLEB1P5, mRNA sequence.

ACCESSION AM036063 GI:5894905
 VERSION AM036063.1
 KEYWORDS EST.
 SOURCE
 ORGANISM Lycopersicon esculentum
 tomato.

REFERENCE
 1 (bases 1 to 480)
 Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Viston, T., Holt, J.E., Liang, F., Upton, J., Romning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nieman, M., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
 Generation of ESTs from tomato seed tissue
 Unpublished (1999)
 CONTACT: CUGI
 CLEMSON UNIVERSITY Genomics Institute
 CLEMSON UNIVERSITY
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES
 source
 1. 480
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="CLEB1P5"
 /clone_1lb="tomato seed, TAMU"
 /tissue-type="seeds"
 /dev_stage="quiescent seed"
 /lab_host="XLI-Blue MRF"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; CLEB - Tomato Seed EST Library. Directionally cloned cDNAs inserted into pBluescript SK(-) at 5' end with EcoRI and 3' end with XhoI site."

BASE COUNT 175 a 79 c 64 g 162 t
 ORIGIN

alignment_scores:
 Quality: 75.00 Length: 80
 Ratio: 1.705 Gaps: 5
 Percent Similarity: 55.000 Percent Identity: 31.250

alignment_block:
 US-09-728-670-10 x AM036063 ..

Align seg 1/1 to: AM036063 from: 1 to: 480

```

69 AspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProSerAl 85
|||||:|||||:|||||:|||||:|||||:|||||:
138 GATGAAGACGCTTATACACTATTTCAC.....GAACCATCTCC 175
85 alyslleGlu.....ValThrTyrT 92
176 AAAAATGTAAGACGCTTATACACTATTTCAGAAATCATCTCGAGATATT 225
92 yraspLysAsnLysLysLysGlu.....GluThrLysSerPheProIle 106
|||||:|||||:|||||:|||||:|||||:|||||:
226 ACGATGAATAATGAAAGAGGTTTCAAAAGAGTAATTCGTTTATATG 275
107 ThrGluLysGlyPheValValProasp..LeuSerGluHisIleLysAsn 123
|||||:|||||:|||||:|||||:|||||:|||||:
276 CTACCAAAAGAGGAGCTTAATCTCTCTTGAGCACCATGACGTAACACA 324
123 roGlyPheAsnLeuIleThrLysValValIleGluLys 135
325 .....ATGAGATGTGACTTATCTAATTAATTCCTTAAA 357

```

seq_name: gb_est1:AV052970

seq_documentation_block:


```

216 AATTTTCATCTCAATTGTGATCCGGAGAGAGATACGTCAATTCAGT 167
131 lval 132
166 GTTG 163

seq_name: gb_est2:BI207927

seq_documentation_block:
LOCUS      BI207927              522 bp    mRNA    linear    EST 11-JUL-2001
DEFINITION EST529677 cTOS Lycopersicon esculentum cDNA clone cTOS15D23 5' end,
            mRNA sequence.
ACCESSION  BI207927
VERSION    BI207927.1  GI:14685651
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE  1 (bases 1 to 522)
AUTHORS   van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R.,
            Ronning, C. and Tanksley, S.
TITLE      Generation of ESTs from Tomato Suspension Cultures
JOURNAL    Unpublished (2001)
COMMENT    Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
            Location/Qualifiers
                source          1..522
                                /organism="Lycopersicon esculentum"
                                /cultivar="TA96, E6203"
                                /db_xref="taxon:4081"
                                /clone="cTOS15D23"
                                /clone_1lb="cTOS"
                                /tissue_type="suspension cultures"
                                /lab_host="SOLR"
                                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                                XhoI; Suspension cultures of L.esculentum E6203 were grown
                                in Murashige and Skoog based medium, supplemented with 15%
                                coconut milk (filter sterilized and added after
                                autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
                                Fresh medium was added every 7 days, and cultures were
                                grown at 25 C, with 12hrs of light and continuous
                                shaking."
BASE COUNT      156 a      103 c      98 g      165 t
ORIGIN
alignment_scores:
    Quality:      74.50      Length:      118
    Ratio:        1.183      Gaps:        5
    Percent Similarity: 53.390      Percent Identity: 24.576

alignment_block:
US-09-728-670-10 x BI207927/rev ..

Align seg 1/1 to reverse of: BI207927 from: 1 to: 522

38 GULenLeuSerProHisTyrValGluPheProIleLysProGlyThrH 54
498 GATTCTCGAACCCTACAGTA.....GGAAGCAG 467
54 rLeuThrLysGluLysIleGluTyrValGluTyrPalalaLeuAspAla 71
466 GATCAAAAACCTCAAAATGATGATACATCACTATGCAGCTGACAGCA 417
71 hrala.....TyrLysGluPheArg 77
:::|

```

```

416 GTGCAAAAATGAACCCGTGGGAGAGTTTATGATAGAAGTCCGAGAGA 367
78 ValValGluLeuasp.....ProSerAlaLysIleGluValThrTy 91
366 GTGATGAGTGTAAAGCATCTGCACCTGGGTTCATTAATTCACATTCOA 317
91 rTyrAspLysAsnLysLysLysGluLutThrLysSerPhePro..... 105
316 TTGGGATTAAGTAAAGGAAAGGTGATTTATGATATCCGCCATCAG 267
106 ....lLeThrGluLysGlyPheValValProAspLeuSerGluHis... 119
266 CACCATTTGGCTTTGGAACCTCTAGTATTTCCAGACCGCTGAAATCACCC 217
120 .....lLeLysAsnProGlyPheAsnLeuLeuThrLysVa 131
216 AATTTTCATCTCAATTGTGATCCGGAGAGAGATACGTCAATTCAGT 167
131 lval 132
166 GTTG 163

seq_name: gb_est2:BI211112

seq_documentation_block:
LOCUS      BI211112              522 bp    mRNA    linear    EST 11-JUL-2001
DEFINITION EST529152 cTOS Lycopersicon esculentum cDNA clone cTOS25A10 5' end,
            mRNA sequence.
ACCESSION  BI211112
VERSION    BI211112.1  GI:14688836
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE  1 (bases 1 to 522)
AUTHORS   van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R.,
            Ronning, C. and Tanksley, S.
TITLE      Generation of ESTs from Tomato Suspension Cultures
JOURNAL    Unpublished (2001)
COMMENT    Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
            Location/Qualifiers
                source          1..522
                                /organism="Lycopersicon esculentum"
                                /cultivar="TA96, E6203"
                                /db_xref="taxon:4081"
                                /clone="cTOS25A10"
                                /clone_1lb="cTOS"
                                /tissue_type="suspension cultures"
                                /lab_host="SOLR"
                                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                                XhoI; Suspension cultures of L.esculentum E6203 were grown
                                in Murashige and Skoog based medium, supplemented with 15%
                                coconut milk (filter sterilized and added after
                                autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
                                Fresh medium was added every 7 days, and cultures were
                                grown at 25 C, with 12hrs of light and continuous
                                shaking."
BASE COUNT      155 a      103 c      99 g      165 t
ORIGIN
alignment_scores:
    Quality:      74.50      Length:      118
    Ratio:        1.183      Gaps:        5
    Percent Similarity: 53.390      Percent Identity: 24.576

```

alignment_block:

US-09-728-670-10 x B1211112/rev ..

Align seg 1/1 to reverse of: B1211112 from: 1 to: 522

```

38 GluleuleuserProH1sTyrrValGlupheProIlelyeProGlythr 54
   ::::::::::::::::::::
498 GATTTCCTGAACCCATACGTA.....GGAAAGCAG 467
54 rleuthrlyglulysileglutrryrrValglutrralaleuaspalat 71
   ::::::::::::::::::::
466 GATCAAAAAAATCTCAAAATGATATGATCATCACTATGATCGACAGCA 417
71 hrAla.....Tyrlysglupearg 77
   ::::::::::::::::::::
416 GTCCAAATAATGAAACCCGTGGAGAGATTATGATTAAGATCCGCAAGA 367
78 ValValgluleuasp.....ProserAlalysilegluValthr 91
   ::::::::::::::::::::
366 GTGATGTAGTGAAGCATCTGCACCTGGTTCATTTCTACCTTCGAA 317
91 rTyrrAspLysAsnLysLysLysgluLutrrLysSerPhePro..... 105
   ::::::::::::::::::::
316 TTGGGATTAAGTAAAGGAAAGGTGATTTATGATCCGCCCTCAG 267
106 .....lIethrGluysGlyPheValValProAspleuserGluHs... 119
   ::::::::::::::::::::
266 CACCATTCGCTTGGAAACCTAGTATTTCCAGACGCTGTGATCACCC 217
120 .....lIelysAsnProGlyPheAsnleulelthrlysg 131
216 AATTTCCTCATCTCATATTGTGAATCGGGAGAGAGATCATTCATTCAG 167
131 lVal 132
   ::::::::::::::::::::
166 GTTG 163

```

seq_name: gb_est2:B1955134

seq_documentation_block:

LOCUS B1955134 679 bp mRNA linear EST 19-Oct-2001
 DEFINITION HVSEMEM002119f Hordeum vulgare green seedling EST library
 HVCNDA0014 (Blumeria infected) Hordeum vulgare cDNA clone
 HVSEMEM002119f, mRNA sequence.

ACCESSION B1955134
 VERSION B1955134 GI:16301210
 KEYWORDS EST.

SOURCE barley.
 ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 679)
 Wing,R., Close,T.J., Kleinof,A., Wise,R., Chin,A., Begum,D.,
 Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons
 ,J., Oates,R. and Main,D.

TITLE Development of a genetically and physically anchored EST resource
 for barley genomics: Blumeria infected Morex (compatible) seedling
 cDNA library
 JOURNAL Unpublished (2001)
 COMMENT Contact: Ming RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 479
 Seq primer: AATTAACTCTCACTAAAGG
 High quality sequence stop: 657.
 Location/Qualifiers
 1..679
 /organism="Hordeum vulgare"

FEATURES
 Source

/cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSEMEM002119f"
 /clone_1b="Hordeum vulgare green seedling EST library
 HVCNDA0014 (Blumeria infected)"
 /issue_type="green seedling leaf"
 /lab_host="TUC121"

/note="Vector: Bluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; Morex (mla) plants were greenhouse grown in the R
 Wise lab at Iowa State University, Ames, IA; 7 day old
 green seedlings were infected with isolate 5874 of
 Blumeria graminis f. sp. hordei, and leaves were harvested
 24, 48 and 72 hr post-inoculation and snap frozen (Wise).
 In the TUC121 lab at the University of California,
 Riverside, total RNA was prepared from each sample pool,
 equal quantities of all three RNA pools were combined,
 poly(A) RNA was purified from the mixture, one primary
 unamplified cDNA library was made, and 1 million pfu were
 in vivo excised to give bluescript SK(-) cDNA phagemids
 (Chin). Phagemids were plated and picked at the Clemson
 University Genomics Institute (CUGI) (Begum, Palmer,
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
 sequencing and sequence analysis were performed at CUGI
 (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
 The sequence has been trimmed to remove vector sequence
 and contains a minimum of 100 bases of phred value 20 or
 above. For more details on library preparation and
 sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
 this clone see <http://www.genome.clemson.edu/orders> Also
 see Close TJ, Wing R, Kleinof A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (<http://wheat.pw.usda.gov/gspages/bgn/31/cover.html>)

BASE COUNT 188 a 121 c 178 g 189 t 3 others
 ORIGIN

alignment_scores:
 Quality: 74.50 Length: 99
 Ratio: 1.552 Gaps: 3
 Percent Similarity: 48.485 Percent Identity: 25.253

alignment_block:
 US-09-728-670-10 x B1955134 ..

Align seg 1/1 to: B1955134 from: 1 to: 679

```

33 AspsrlysglyAsnGluleuleuser..... 41
   ::::::::::::::::::::
170 GACTCTGAAGGAGCTAAACCTTATGAGCGGTTGAGTCACGGTAT 219
42 .ProH1sTyrrValGlupheProIlelyeProGlythrIethrlysg 58
   ::::::::::::::::::::
220 TCCACACCTGTTTCTTCCTGATGACAAAACGTGTAATCTTCTATGATG 269
58 lulsileglutrryrrValglutrralaleuaspalThrAlatrrlys 74
   ::::::::::::::::::::
270 AAGGAGTTGAGTTTGTAGTGAATATGATGAGACTTATTCCTTTACA 319
75 GlupheargValValgluleuaspProserAlalysilegluValthr 91
   ::::::::::::::::::::
320 ACTGAGAGCATCAATGATTAAGGAGCA..... 349
91 rTyrrAspLysAsnLysLysLysgluLutrr..... 101
   ::::::::::::::::::::
350 ....GAAAGGACGCTAAGGATATCAAACTTATGATGCTGCTGTA 395
102 .....LysSerPheProIlethrGluysGlyPheValValPro 114
   ::::::::::::::::::::
396 CAGCAAAATCTGCTCAAGTAATTCAAACACGGGAGAGAGGTACCC 442

```

seq_name: gb_est2:B1955134


```

seq_documentation_block:
  LOCUS       BF570018                993 bp      mRNA       linear      EST 12-DEC-2000
  DEFINITION   602186015T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4310571 3',
  ACCESSION   BF570018
  VERSION     BF570018.1  GI:11643730
  KEYWORDS    EST.
  SOURCE      human.
  ORGANISM    Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1 (bases 1 to 993)
               NIH-MGC http://mgi.nci.nih.gov/.
               National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE       Unpublished (1999)
  JOURNAL     Contact: Robert Strausberg, Ph.D.
               Email: cga@bbs-remail.nih.gov
               Tissue Procurement: Lineman
               CDNA Library Preparation: Ling Hong/Rubin Laboratory
               CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LICM185 row: m column: 04
               High quality sequence start: 14
               High quality sequence stop: 714.
  FEATURES    Location/Qualifiers
               source          1..993
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:4310571"
               /clone_1ib="NIH_MGC_45"
               /tissue_type="renal carcinoma (ascites)"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: kidney; Vector: pOTB1; site_1: XhoI; site_2:
               EcoRI; cDNA made by oligo-dT priming. Directionally cloned
               into EcoRI/XhoI sites using the following 5' adaptor:
               GGCACGAG(G). Library constructed by Ling Hong in the
               laboratory of Gerald M. Rubin (University of California,
               Berkeley) using Zap-cDNA synthesis kit (Stratagene) and
               Superscript II RT (Life Technologies). Note: this is a
               NIH-MGC library."
  BASE COUNT   170 a      293 c      256 g      274 t
  ORIGIN
  alignment_scores:
    Quality: 74.50      Length: 114
    Ratio: 1.284      Gaps: 7
    Percent Similarity: 50.877      Percent Identity: 30.702
  alignment_block:
    US-09-728-670-10 x BF570018/rev ..
  Align seg 1/1 to reverse of: BF570018 from: 1 to: 993
  20 Prothrglyprotyrleumetvalasnvalthr..... 30
  ||| ||||| ||| : : : : : |||
  796 CCTTGGGACCAACGCTGTGGCAGGCAACCTCCAGGTGAGAGACT 747
  ||| : : : : : |||
  31 .GlyValAspSerIysGlyAsnGluLeuLeuSerProHisIstyrValGluP 47
  ||| : : : : : ||| : : : : : ||| : : : : : |||
  746 GGGAAACCGCTTATGAGGGAAGCGGCTCTGGAACCAACATCTGTGAT 697
  47 heproilleysproglythrthrleuthrlysglu.lysilleglutyrty 63
  ||| : : : : : |||
  696 TTCAGACCAAG.....ACTGAGGAGAAG..... 673
  ||| : : : : : |||
  63 rValGluIrpAlaLeuAspAlaThrAlaIstyrlysgluIupheargValValG 80
  ||| ||||| ||| : : : : : |||
  672 .....ACACTTGATCGGACCCCTTTGGGGG.....G 648

```

```

seq_documentation_block:
  LOCUS       AW036324                483 bp      mRNA       linear      EST 18-MAY-2001
  DEFINITION   EST778449 tomato seed, TAMU Lycopersicon esculentum cDNA clone
  ACCESSION   AW036324
  VERSION     AW036324.1  GI:5895078
  KEYWORDS    EST.
  SOURCE      tomato.
  ORGANISM    Lycopersicon esculentum
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
               Lycopersicon.
  REFERENCE   1 (bases 1 to 483)
               Alcalá,J., Vredalov,J., White,R., Matern,A.L., Viston,T., Holt,I.E.,
               Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
               ,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
               ,S.D. and Giovannoni,J.
               Generation of ESTs from tomato seed tissue
  TITLE       Unpublished (1999)
  JOURNAL     Contact: CUGI
               Clemson University Genomics Institute
               Clemson University
               100 Jordan Hall, Clemson, SC 29634, USA
               Email: http://www.genome.clemson.edu/orders/index.html
  FEATURES    Location/Qualifiers
               source          1..483
               /organism="Lycopersicon esculentum"
               /cultivar="TA496"
               /db_xref="taxon:4081"
               /clone="CLEB1J4"
               /clone_1ib="tomato seed, TAMU"
               /tissue_type="seeds"
               /dev_stage="quiescent seed"
               /lab_host="XLI-Blue MRF"
               /note="Vector: pBluescript SK(-); site_1: EcoRI; site_2:
               XhoI; cDE - Tomato Seed EST Library. Directionally cloned
               cDNAs inserted into pBluescript SK(-) at 5' end with
               EcoRI and 3' end with XhoI site."
  BASE COUNT   179 a      77 c      64 g      163 t
  ORIGIN
  alignment_scores:
    Quality: 74.00      Length: 80
    Ratio: 1.762      Gaps: 5
    Percent Similarity: 52.500      Percent Identity: 32.500
  alignment_block:
    US-09-728-670-10 x AW036324 ..
  Align seg 1/1 to: AW036324 from: 1 to: 483
  69 AsplArthArIatYrlysgluIupheargValValGluLeuAspProseral 85
  ||| ||||| ||| : : : : : |||
  138 GATGAAGACGCTTATACACTATTTCAC.....GACCAATCTCC 175

```

```

85 alysiileglu.....ValThrTyrT 92
176 AAAATATGAGACGCTTATACACTATTTCAGATCATCTCCGAGATATT 225
92 yraspPlysaNlyslYslYsgluThrlYsser.....PheProile 106
226 ACATATGAAATGAAAGAAAGACTTCAAAAGAGTAATTGTTTCATATG 275
107 ThrGluYslGlyPheValValProasp.LeuSerGluHisIleYsaSnp 123
276 CTACCAAAAGGAGCTTAATTCCTCTTGACACATCAGTAGACACA. 324
123 roGlyPheasNleuIleThrlYslValValIleGluYls 135
325 .....ATGGAGATGGTACTTATCTAATTAATCTTAA 357

seq_name: qb_esc1:AM036351

seq_documentation_block:
LOCUS AM036351 488 bp mRNA linear EST 18-MAY-2001
DEFINITION EST278476 tomato seed, TAMU Lycopersicon esculentum cDNA clone
CLEE1K4, mRNA sequence.
ACCESSION AM036351
VERSION AM036351.1 GI:5895105
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; eumasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 488)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vlsion,T., Holt,I.E.,
Liang,F., Upton,J., Romling,C.M., Craven,M.B., Fujii,C.Y., Bowman
,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
,S.D. and Giovannoni,J.
Generation of ESTs from tomato seed tissue
Unpublished (1999)
JOURNAL Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Source Location/Qualifiers
1..488
/organism="Lycopersicon esculentum"
/cultivar="TAA96"
/db_xref="taxon:4081"
/clone_lib="tomato seed, TAMU"
/tissue_type="seeds"
/dev_stage="quiescent seed"
/lab_host="XLI-Blue MRP"
/notes="Vector: pBluescript SK(-); Site1: EcoRI; Site2:
XhoI; cDNA - Tomato Seed EST library. Directionally cloned
cDNAs inserted into pBluescript SK(-) at 5' end with
EcoRI and 3' end with XhoI site."
BASE COUNT 180 a 78 c 64 g 163 t 3 others
ORIGIN
alignment_scores:
Quality: 74.00 Length: 80
Ratio: 1.762 Gaps: 5
Percent Similarity: 52.500 Percent Identity: 32.500

alignment_block:
US-09-728-670-10 x AM036351 ..
Align seq 1/1 to: AM036351 from: 1 to: 488
69 AspaletThrAlaTyrTrpYsgluPheAlcValValGluLeuAspProSerAl 85

```

```

138 GATGAGACGCTTATACACTATTTC.....GACCATCTCC 175
85 alysiileglu.....ValThrTyrT 92
176 AAAATATGAGACGCTTATACACTATTTCAGATCATCTCCGAGATATT 225
92 yraspPlysaNlyslYslYsgluThrlYsser.....PheProile 106
226 ACATATGAAATGAAAGAAAGACTTCAAAAGAGTAATTGTTTCATATG 275
107 ThrGluYslGlyPheValValProasp.LeuSerGluHisIleYsaSnp 123
276 CTACCAAAAGGAGCTTAATTCCTCTTGACACATCAGTAGACACA. 324
123 roGlyPheasNleuIleThrlYslValValIleGluYls 135
325 .....ATGGAGATGGTACTTATCTAATTAATCTTAA 357

seq_name: qb_esc2:BI391530

seq_documentation_block:
LOCUS BI391530 613 bp mRNA linear EST 06-AUG-2001
DEFINITION pgpin.pk004.119 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library Gallus gallus cDNA clone pgpin.pk004.119 5' similar to
gb|AAK51065.1|AF361746.1 (AF361746) endochelital cell-selective
adhesion molecule [Homo sapiens], mRNA sequence.
ACCESSION BI391530
VERSION BI391530.1 GI:15084812
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 613)
Porter,T.E. and Cogburn,L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
library USDA/IRAFs Animal Genome Project
Unpublished (2001)
JOURNAL Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES
Source Location/Qualifiers
1..613
/organism="Gallus gallus"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone_lib="pgpin.pk004.119"
/clone_lib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (dl2,dl4,dl9); post-hatch (w1,w3,w5
,w7,w9)"
/lab_host="E. coli EMP10B"
/notes="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"
BASE COUNT 118 a 197 c 183 g 99 t 16 others
ORIGIN
alignment_scores:
Quality: 74.00 Length: 110
Ratio: 1.370 Gaps: 4
Percent Similarity: 49.091 Percent Identity: 26.364

alignment_block:
US-09-728-670-10 x BI391530 ..

```

LOCUS	654 bp	mRNA	linear	EST 15-JUN-2001
DEFINITION	pf1in.pk005.k14 normalized chicken fat cDNA library Gallus gallus			
DEFINITION	cDNA clone pf1in.pk005.k14 5' similar to g1110048456			
DEFINITION	ref NP_065266.1 cortical thymocyte receptor (X. laevis CTX) like;			
DEFINITION	CTM gene [Mus musculus] gb A0157524.1 (Af061024) CTM [Mus musculus]			
DEFINITION	jG. mRNA sequence.			
ACCESSION	BI065721			
VERSION	BI065721.1	GI:14473243		
KEYWORDS	EST.			
SOURCE	chicken.			
ORGANISM	Gallus gallus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Archosauvia; Aves; Neognathae; Galliformes; Phasianidae;			
	Phasianinae; Gallus.			
	1 (bases 1 to 654)			
REFERENCE	Cogburn,L.A., Morgan,R.W. and Burnside,J.			
AUTHORS	Chicken ESTs from fat			
TITLE	Unpublished (2001)			
JOURNAL	Contact: Larry A. Cogburn			
COMMENT	University of Delaware			
	Townsend Hall, Newark, DE 19717, USA			
	Tel: 302-831-1335			
	Fax: 302-831-2822			
	Email: cogburn@udel.edu, www.chickest.udel.edu.			
FEATURES	Location/Qualifiers			
SOURCE	1..654			
	/organism="Gallus gallus"			
	/db_xref="taxon:9031"			
	/clone="pf1in.pk005.k14"			
	/clone_lib="normalized chicken fat cDNA library"			
	/sex="Male and Female"			
	/tissue_type="fat"			
	/lab_host="E.coli EMDH10B"			
	/note="Vector: pSPORT1"			
BASE COUNT	133 a	206 c	193 g	111 t
ORIGIN	11 others			
alignment_scores:				

```

seq_name: gb_est2:BI929102
seq_documentation_block:
  LOCUS      718 bp      mRNA      linear      EST 18-OCT-2001
  DEFINITION  EST548991 tomato flower, 3 - 8 mm buds Lycopersicon esculentum CDNNH
  ACCESSION  clone CTOB27B22 5' end, mRNA sequence.
  KEYWORDS   BI929102
  SOURCE     BI929102.1  GI:16242002
  ORGANISM   EST.
             tomato.
             Lycopersicon esculentum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
             Lycopersicon.
  REFERENCE  1 (bases 1 to 718)
             van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
             Mittlebach,T., Gao Aken,S., Ronning,C.M., Nieman,W., Frazer,C.M.,
             Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
             Generation of ESTs from tomato flower tissue, 3-8 mm buds (2001)
             Unpublished (2001)
  TITLE      Contact: CUGI
  JOURNAL    Clemson University Genomics Institute
  COMMENT    Clemson University
             100 Jordan Hall, Clemson, SC 29634, USA
             Email: http://www.genome.clemson.edu/orders/index.html
             This clone is available through the Clemson University Genomics
             Institute
  Seq primer: T3.
             Location/Qualifiers
             1..718
             /organism="Lycopersicon esculentum"
             /cultivar="TA496"
             /db_xref="taxon:4081"
             /clone="CTOB27B22"
             /clone_1lb="tomato flower, 3 - 8 mm buds"

```



```

SOURCE      Populus tremula x Populus tremuloides.
ORGANISM    Populus tremula x Populus tremuloides.
REFERENCE   1 (bases 1 to 355)
AUTHORS     Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A., Holmberg
            'A., Amlid,B., Bhalerao,R., Larsson,M., Villarroel,R., Van Montagu
            'M., Sandberg,G., Olsson,O., Teerl,T.T., Boerjan,W., Gustafsson,P.,
            Uhlen,M., Sundberg,B. and Lundberg,J.
            Gene discovery in the wood-forming tissues of poplar: Analysis of
            5,692 expressed sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
TITLE       Contact: Sterky F
JOURNAL     Department of Biotechnology
MEDLINE     Royal Institute of Technology (KTH)
COMMENT     Teknikringen 34, S-100 44 STOCKHOLM, Sweden
            Tel: +46 8 790 8287
            Fax: +46 8 24 54 52
            Email: fredrik@biochem.kth.se
            PCR Primers
            FORWARD: AAAGGGGAGTGTCTGCAAGCGC
            BACKWARD: GCTTCGGCGTCGTATGTTGTGTG
            Seq primer: CGTGTAAACGACGCGCAG
            High quality sequence stop: 355.
            Location/Qualifiers
                1..355
                /organism="Populus tremula x Populus tremuloides"
                /db_xref="taxon:47664"
                /clone_lib="Hybrid aspen plasmid library"
                /tissue_type="Cambial region"
                /dev_stage="1.5 m actively growing tree"
                /lab_host="E.coli"
                /note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI
                /note="Cambial region tissues, including developing xylem, the
                /note="meristematic cambial zone and the developing and mature
                /note="phloem, was harvested from 1.5 m actively growing trees.
                /note="cDNA was prepared and cloned into lambda gt2a. DNA was
                /note="isolated and subcloned into pBluescript SK using SalI and
                /note="NotI restriction enzymes."
BASE COUNT  118 a      79 c      87 g      68 t      3 others
ORIGIN
alignment_scores:
    Quality: 73.00      Length: 56
    Ratio: 2.281      Gaps: 1
    Percent Similarity: 57.143      Percent Identity: 30.357
alignment_block:
US-09-728-670-10 x A1164607 ..
Align seg 1/1 to: A1164607 from: 1 to: 355
2 serSerPheasplsglyTyrTylsylvsglyAspaspalaasertyrph 18
||||| ::|||::| ::|||::| ::|||::|
35 TCAAGTCGAGGAAGGCAAAACACGAGAGNGGGAANCAAGAGCATTTT 84
18 egluprOthglyPro.....TyrL 25
||||| ::|||
85 CGAGGCAATGTCTCTMAAGCCTCACTTAAGCCTCAACGACGACGATTT 134
25 euuetValasnValThrcglyValaspserysglyAanglunleuser 41
::|::| ::|||::| ::|||::| ::|||::|
135 CCATTCATATACGTTGATGCGCTCCCACTAAGGCGATCAAAAGCTTACC 184
42 ProHistyValgluphe 47
||||| ::|||::| ::|||::|
185 CCAGAAATATGTTGACTTT 202
seq_name: gb_est2:T19018

```

```

Seq. documentation block:
LOCUS       T19018                402 bp    mRNA    linear    EST-25-SEP-1996
DEFINITION  f11010t Testis 1 Homo sapiens cDNA clone f11010 5' end, mRNA
sequence.
ACCESSION   T19018
VERSION     T19018.1
KEYWORDS    GI:601061
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 402)
AUTHORS     Pavlak, A., Toussaint, C., Levy, I., Bulle, F., Poyard, M., Barouki, R.
            and Guellien, G.
TITLE       Characterization of a large population of mRNAs from human testis
JOURNAL     Genomics 26, 151-158 (1995)
MEDLINE     95301283
COMMENT     Contact: Guellien G
            Unite INSERM 99
            INSERM
            Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
            Tel: (33)149813530
            Fax: (33)1498980908
            Email: guellien@infobiogen.fr
FEATURES
     source          1..402
                     Location/Qualifiers
                         1..402
                         /organism="Homo sapiens"
                         /db_xref="taxon:9606"
                         /clone="f11010"
                         /clone_1lb="Testis 1"
                         /note="Vector: pGEM 5zf(+); Site_1: EcoRV; Site_2: NotI;
                         cDNA was prepared from human testis of a 27 years old man.
                         cDNA was prepared using a 15mer oligo dT anchored by two
                         degenerated bases at its 3' end and containing a NotI site
                         at its 5' end. The cDNA was cloned between EcoRV and NotI
                         sites of pGEM 5zf(+). The 3' end is at the NotI site. The
                         EcoRV site is lost during the cloning procedure. cDNA
                         corresponding to abundant species were eliminated from
                         this library."
BASE COUNT      130 a          66 c          97 g          102 t          7 others
ORIGIN
alignment_scores:
    Quality:      73.00          Length:      115
    Ratio:        1.237          Gaps:      5
    Percent Similarity: 51.304    Percent Identity: 27.826
alignment_block:
US-09-728-670-10 x T19018 ..
Align seg 1/1 to: T19018 from: 1 to: 402
26 MetValAsnValThrGlyVal...AspSerLygLyAsnGluLeuLeuSe 41
   :::::::::::::::::::: ||| ||| :::::::::::::: ::
17 ATAGTTGACATCTACTTTTGGATGACATAAGAGGAGCGCAAAAACAGC 66
   : |||||||:|||||:|||||: ||| :|||:|||||:
41 rProHLSrTyValGluPhe.....ProIleLysProGluYThrThreIue 56
   : |||||||:|||||:|||||: ||| :|||:|||||:
67 GGAAGCCCTATGTGCAATTTGAAGAACCCAGAAATGGCCCAAGCCCTGT 116
56 hrLygSguLySILegIuTYrTYrValGluTrpAlaLeuAspAlaThrAla 72
   |||:|||||: |||
117 TGAAGACGACGAGGAAGAAAT..... 136
73 TyrLygSguLpheArGyValValGluLeuAspProSerAlaLySILegIuVa 89
   ||| ::::::::::: |||||| ::|
137 .....GGTAATCGATACATCGAGATATTTCCAAAGCAGAGATGAACT 180
89 lThrTYrTYrAspLyAsnLySILySguGluGluThrLySerPheProI 106
   |::::::::: ::|
181 TCGAACACATGTCGGTCTTATTAAGGGAAGAAAGAAATCGACTTTTCTCT. 229

```


